

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 17, 2003, 09:13:37 ; Search time 3691 Seconds  
(without alignments)  
10496.184 Million cell updates/sec

Title: US-10-043-539A-1

Perfect score: 947

Sequence: 1 gtttcaaatcgggtgagg.....gtttcaaatcgaatcgat 947

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

GenEmbl.\*

1: gb\_ba.\*

2: gb\_hgt.\*

3: gb\_in.\*

4: gb\_on.\*

5: gb\_ov.\*

6: gb\_pat.\*

7: gb\_ph.\*

8: gb\_pl.\*

9: gb\_pr.\*

10: gb\_ro.\*

11: gb\_sts.\*

12: gb\_sy.\*

13: gb\_un.\*

14: gb\_vl.\*

15: em\_ba.\*

16: em\_fun.\*

17: em\_hum.\*

18: em\_in.\*

19: em\_ru.\*

20: em\_on.\*

21: em\_or.\*

22: em\_ov.\*

23: em\_pat.\*

24: em\_ph.\*

25: em\_pl.\*

26: em\_ro.\*

27: em\_sts.\*

28: em\_un.\*

29: em\_vl.\*

30: em\_hgt\_hum.\*

31: em\_hgt\_inv.\*

32: em\_hgt\_other.\*

33: em\_hgt\_mus.\*

34: em\_hgt\_pln.\*

35: em\_hgt\_rtd.\*

36: em\_hgt\_mam.\*

37: em\_hgt\_vrt.\*

38: em\_sy.\*

39: em\_hgt\_hum.\*

40: em\_hgt\_mus.\*

41: em\_hgt\_other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match %	Length	DB	ID	Description
1	947	100.0	947	1	AF207701	Staphyloc
C 2	826.4	87.3	304050	1	AF004829	Staphyloc
C 3	826.4	87.3	307750	1	AP003136	Staphyloc
C 4	824.8	87.1	348650	1	AP003364	Staphyloc
C 5	369.8	39.0	384	6	AX620460	Sequence
C 6	307	32.4	306898	1	AE016750	Staphyloc
C 7	199.4	21.1	834	6	AX620416	Sequence
C 8	92.8	10.2	8056	6	AX599046	Sequence
C 9	92.8	9.8	170627	2	AC125567	Rattus no
C 10	91.4	9.7	67970	3	PFMALIP3	AL031746 Plasmodiu
C 11	90.6	9.6	14422	3	AF46146	Melipona
C 12	90.4	9.5	86826	3	PFMAL3P5	AL034556 Plasmodiu
C 13	90	9.5	8056	6	AX599046	Sequence
C 14	87.6	9.3	258658	3	AE014832	Plasmodiu
C 15	87.2	9.2	8056	6	AX598900	Sequence
C 16	85.2	9.0	286208	2	AC117140	Rattus no
C 17	85	9.0	840	8	CNS018QK	Botrytis
C 18	84.8	9.0	2009	6	AX457067	Sequence
C 19	84.6	8.9	250029	3	AE014820	Plasmodiu
C 20	84.2	8.9	259474	9	HUAC004605	Homo sapi
C 21	83.6	8.8	175544	2	AC117342	Rattus no
C 22	83	8.7	1090	3	AB084761	AC092086 Homo sapi
C 23	82.8	8.7	169046	5	AL928828	zebrafish
C 24	82.8	8.7	348600	1	AB063521	Wigglesw
C 25	82.4	8.7	349980	6	AX344555	Sequence
C 26	82.4	8.7	196875	2	AC098095	Sequence
C 27	82.2	8.7	6292	6	AX251433	Sequence
C 28	82	8.7	233877	9	AC093798	Homo sapi
C 29	82	8.6	349751	3	PFMAL4P3	AC035476 Plasmodiu
C 30	81.6	8.6	156060	2	AC004153	Plasmodiu
C 31	81.4	8.6	250707	3	AE014848	Plasmodiu
C 32	81.4	8.6	347050	3	PEA929351	Plasmodiu
C 33	81.4	8.6	99003	2	AL390756	Homo sapi
C 34	81.2	8.6	110000	2	PFMAL3_09	Continuation (10
C 35	81.2	8.6	318221	2	PFMAL13P3	AL049184 Plasmodiu
C 36	81.2	8.6	162445	9	AL158151	Human DNA
C 37	81	8.6	164640	2	BX005461	BX005461 Danio rer
C 38	81	8.6	168558	2	BX322549	Danio rer
C 39	81	8.6	84563	5	AC096885	Danio rer
C 40	80.8	8.5	313050	3	PEA929352	Plasmodiu
C 41	80.6	8.5	8056	6	AX598900	Sequence
C 42	80.2	8.5	60604	2	AC023466	Homo sapi
C 43	80.2	8.5	181792	9	AC098822	Homo sapi
C 44	80.2	8.4	112695	8	AC119418	Medicago
C 45	80					

ALIGNMENTS

RESULT 1	AF207701	947 bp	DNA	linear	BCT 01-DEC-2000
LOCUS	AF207701	947 bp	DNA	linear	BCT 01-DEC-2000
DEFINITION	Staphylococcus aureus putative transposase gene, partial cds; and				
ACCESSION	AF207701				
VERSION	AF207701.1				
KEYWORDS	GI:11493941				
SOURCE	Staphylococcus aureus				
ORGANISM	Staphylococcus aureus				
REFERENCE	1. (bases 1 to 947)				
AUTHORS	Cheung, A.L. and Manna, A.C.				
TITLE	Characterization of sarR, a modulator of sar expression in				
	Staphylococcus aureus				

JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 947)  
 AUTHORS Cheung,A.L. and Manna,A.C.  
 TITLE Direct Submission  
 JOURNAL Submitted (22-NOV-1999) Microbiology, Dartmouth Medical School,  
 College St., Vail 206, Hanover, NH 03755, USA  
 FEATURES Location/Qualifiers  
 source 1..947  
 /organism="Staphylococcus aureus"  
 /mol\_type="genomic DNA"  
 /strain="RN6390"  
 /db\_xref="taxon:1280"  
 <1..70  
 /codon\_start=2  
 /transl\_table=11  
 /product="putative transposase"  
 /protein\_id="AAG35716.1"  
 /db\_xref="GI:11493943"  
 /translation="FSKSVEVHEKVIHFLKIKYQ"  
 gene 208..555  
 /gene="sarB"  
 /note="regulatory gene"  
 208..555  
 /gene="sarR"  
 /note="similar to SarA"  
 /codon\_start=1  
 /transl\_table=11  
 /product="sarP"  
 /protein\_id="AAG35715.1"  
 /db\_xref="GI:11493942"  
 /translation="MSKINDINDLVNATVQKFFRDTKKKFNLVVEEIIYILNHLRS  
 ESNEISSEKIAKCEFPKPYLTKALQKLDKLSKKRSIQDERTVIVYVDTQKANI  
 OKLISELEXYKN"  
 CDS 374 a 112 c 145 g 316 t  
 BASE COUNT 374 a 112 c 145 g 316 t  
 ORIGIN  
 Query Match 100.0%; Score 947; DB 1; Length 947;  
 Best Local Similarity 100.0%; Pred. No. 7.7e-126;  
 Matches 947; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 1 GTTTTCAAAATCGGTGGAGTGCATGAAAAGTTATTGGGCATTTTGTGAAATAAAAA 60  
 Db 1 GTTTTCAAAATCGGTGGAGTGCATGAAAAGTTATTGGGCATTTTGTGAAATAAAAA 60  
 61 ATATCAATAAGTTCGAGTGCATACCGAATTTTATACCTATTGTTTAGAATGAACCTTA 120  
 Db 61 ATATCAATAAGTTCGAGTGCATACCGAATTTTATACCTATTGTTTAGAATGAACCTTA 120  
 121 TAACATAGTTGGATAGAGTTTCGATTTAATACATTAATGTAACCTGCTACACAAG 180  
 Db 121 TAACATAGTTGGATAGAGTTTCGATTTAATACATTAATGTAACCTGCTACACAAG 180  
 181 ATGTGCATCAGAAGAGTGGTTTAAATAGTAAATTAATGATATTAATGATTTAGTC 240  
 Db 181 ATGTGCATCAGAAGAGTGGTTTAAATAGTAAATTAATGATATTAATGATTTAGTC 240  
 241 AACCCACATTTCAAGTTAGAGTTTTCAGAGATACAAAAGAGTTCATTTTGAAC 300  
 Db 241 AACCCACATTTCAAGTTAGAGTTTTCAGAGATACAAAAGAGTTCATTTTGAAC 300  
 301 TATGAAGAAATTTATATTTTAAATCATATTTTAAAGAGTGAGTCTAACGAAATCTCATCT 360  
 Db 301 TATGAAGAAATTTATATTTTAAATCATATTTTAAAGAGTGAGTCTAACGAAATCTCATCT 360  
 361 AAGAGATTGCTAAGTCTCAGAGTTCAAAACCTTACTATTAACTAAAGCTTTACAAAAG 420  
 Db 361 AAGAGATTGCTAAGTCTCAGAGTTCAAAACCTTACTATTAACTAAAGCTTTACAAAAG 420  
 421 CTAAGAGATTAAATTTGTTATCAAGAAAGAGTTTCAAGACCAAGAGACAGTTATT 480  
 Db 421 CTAAGAGATTAAATTTGTTATCAAGAAAGAGTTTCAAGACCAAGAGACAGTTATT 480  
 481 GTTTATGTTACAGATACAAAAAGCAATATTCAAACAACTGATTTTCAAGATTAGAAGAA 540

Db 481 GTTTATGTTACAGATACAAAAAGCAATATTCAAACAACTGATTTTCAAGATTAGAAGAA 540  
 QY 541 TACATTAAAAATTAATCAAGGTTAATTCGCGTTTAAATACATTGACGATACCAATTTAT 600  
 Db 541 TACATTAAAAATTAATCAAGGTTAATTCGCGTTTAAATACATTGACGATACCAATTTAT 600  
 QY 601 TAATACGAAGTTATTATTTCAGCATTTGGGACATAAAATTAACCTAAAAATTTAAATATTGA 660  
 Db 601 TAATACGAAGTTATTATTTCAGCATTTGGGACATAAAATTAACCTAAAAATTTAAATATTGA 660  
 QY 661 AGATGCTTTTAAATAAGTTAAAGACGACCATACCTTATTTTCAGCTTTTAAAGCTTGACA 720  
 Db 661 AGATGCTTTTAAATAAGTTAAAGACGACCATACCTTATTTTCAGCTTTTAAAGCTTGACA 720  
 QY 721 CAAGGTACACTAGTCTTTTATTATTTTAAATATTTCTTAGAAATCAAGTTTACGATCATAA 780  
 Db 721 CAAGGTACACTAGTCTTTTATTATTTTAAATATTTCTTAGAAATCAAGTTTACGATCATAA 780  
 QY 781 ATATTTTCTCGCATATAGCTTTGGATGGTTCAGATTTTCTCTATATTTTGTGTGCGA 840  
 Db 781 ATATTTTCTCGCATATAGCTTTGGATGGTTCAGATTTTCTCTATATTTTGTGTGCGA 840  
 QY 841 TAAGCAAAAATTTCTAACTGCAAAACCATGTGTAGGCAATTTGAGAAATAGCAACGACAA 900  
 Db 841 TAAGCAAAAATTTCTAACTGCAAAACCATGTGTAGGCAATTTGAGAAATAGCAACGACAA 900  
 QY 901 TCGATGTTATGCTATAAGAACTAATGTTTTCATAAACTGAATCGAT 947  
 Db 901 TCGATGTTATGCTATAAGAACTAATGTTTTCATAAACTGAATCGAT 947  
 RESULT 2  
 AP004829/c 304050 bp DNA linear BCT 20-DEC-2002  
 LOCUS Staphylococcus aureus subsp. aureus MW2, complete genome,  
 strain:MW2, section 8/10.  
 AP004829 BA000033  
 AP004829.1 GI:21205117  
 KEYWORDS  
 SOURCE Staphylococcus aureus subsp. aureus MW2  
 ORGANISM Staphylococcus aureus subsp. aureus MW2  
 Bacteria; Firmicutes; Bacillales; Staphylococcus.  
 REFERENCE 1  
 AUTHORS Baba,T., Takeuchi,F., Kuroda,M., Yuzawa,H., Aoki,K., Oguchi,A.,  
 Nagai,Y., Iwama,N., Asano,K., Naimi,T., Kuroda,H., Cui,L.,  
 Yamamoto,K. and Hiramatsu,K.  
 Genome and virulence determinants of high virulence  
 community-acquired MRSA  
 Lancet 359 (9320), 1819-1827 (2002)  
 22040717  
 PUBMED 12044378  
 REFERENCE 2 (bases 1 to 304050)  
 AUTHORS Director-General, Biotechnology Center, Aoki,K., Oguchi,A., and  
 Nagai,Y., Asano,K., Iwama,N., Baba,T., Kuroda,M., Hiramatsu,K. and  
 Kikuchi,H.  
 TITLE Direct Submission  
 JOURNAL Submitted (06-MAR-2002) Director-General, Biotechnology Center,  
 National Institute of Technology and Evaluation, Biotechnology  
 Center, 2Chome 49-10 Nishihara, Shibuya-ku, Tokyo 151-0066, Japan  
 (E-mail:bio@nites.go.jp, URL:http://www.bio.nite.go.jp/,  
 Tel:81-3-3481-1933, Fax:81-3-3481-8424)  
 Location/Qualifiers  
 1..304050  
 /organism="Staphylococcus aureus subsp. aureus MW2"  
 /mol\_type="genomic DNA"  
 /strain="MW2"  
 /db\_xref="taxon:196620"  
 complement(629..1072)  
 /gene="MW1947"  
 complement(629..1072)  
 /gene="MW1947"  
 /note="ORFID:MW1947"  
 FEATURES  
 source  
 1..304050  
 /organism="Staphylococcus aureus subsp. aureus MW2"  
 /mol\_type="genomic DNA"  
 /strain="MW2"  
 /db\_xref="taxon:196620"  
 complement(629..1072)  
 /gene="MW1947"  
 complement(629..1072)  
 /gene="MW1947"  
 /note="ORFID:MW1947"  
 gene  
 CDS

```

hypothetical protein, similar to acetyltransferase.
[Genomic island nu Sa 4mw]
/codon_start=1
/transl_table=11
/protein_id="BAB95812.1"
/db_xref="GI:21205118"
/transl_table="MSIITRFLNNSPEKLNQCKLYDLDGVTNNDLKKLKITN
HDYFILLIKENKIIGLSGCMQMFYEKNAEYRILAFVIRSEFRKKYGRKLLADS
EFSKRLNCKAITLNSGRNERLSAHLKLYSDNGVSNSTGFTKL"
/complement(1178. .1696)
/gene="MW1948"
CDS
/complement(1178. .1696)
/gene="MW1948"
[Genomic island nu Sa 4mw]
/codon_start=1
/transl_table=11
/product="hypothetical peroteine"
/protein_id="BAB95813.1"
/db_xref="GI:21205119"
/transl_table="MDNINKNEFRVLSTIAITISFILNFGFLISLAIMSIMGKQ
CISUNLLSDTFAVIGWMTLTGAILGAPSTGYGFTSPIDKSSFRASIDWEEQ
RNNVKDMLLIPFTLLSILVIRTFQFASKLENTAIIIFIILVUMFLILLICVK
ITVLCRYFSKE"
/complement(1932. .2522)
/gene="MW1949"
CDS
/complement(1932. .2522)
/gene="MW1949"
[Genomic island nu Sa 4mw]
/codon_start=1
/transl_table=11
/product="truncated hypothetical protein"
/protein_id="BAB95814.1"
/db_xref="GI:21205120"
/transl_table="MSLTAQARFVNEYRTLNVTQSAIKAGYANSAHVATCRLLK
KPHIKQVIOEQDKXIIDENVLTAKELLHLVTATVGETETKEVVKRGYKNPQSG
KVQVLYNEHVELIIVPKPSDRLKAREMLKXHLFDKHDINGNVPIFINIGWDGG
DEGLDKAVKDSNDNPNHVIIVDDIPLEDYESIRFL"
/complement(2519. .2734)
/gene="MW1950"
CDS
/complement(2519. .2734)
/gene="MW1950"
[Genomic island nu Sa 4mw]
/codon_start=1
/transl_table=11
/product="truncated hypothetical protein"
/protein_id="BAB95815.1"
/db_xref="GI:21205121"
/transl_table="WMTDDVNDHIKNHIEWIVDRLATDKFEVIFDLSLIQGLSYQDI
SALDCSEQSVILWYETILDKLVGVK"
/2864. .3338
/gene="truncated int"
/2864. .3211
/gene="truncated int"
[Genomic island nu Sa 4mw]
/codon_start=1
/transl_table=11
/product="truncated int"
/db_xref="GI:21205122"
/transl_table="WKMSDIDPENNTISITYYNPNNNKKFOILPPKTESSTGKIS
VDPNVIKLRDKYINVQNNKRWLNNDHCNEKTINVDTHNVYIYQHTQKLYALIQI
HFLRAHRICTH"
/3072. .3338
/gene="truncated int"
[Genomic island nu Sa 4mw]
/codon_start=1
/transl_table=11
/product="truncated hypothetical protein, similar to integrase"
/protein_id="MW1952"
[Genomic island nu Sa 4mw]
/codon_start=1
/transl_table=11

```

```

/protein_id="BAB95817.1"
/db_xref="GI:21205123"
/transl_table="MITIVMKLSMWIQTIMSRTSILKNISTHFRYTHCALLIESGV
HIKEIOERLRHKDINTMNIYAKITNSYKQDASHKSHQMEDVS"
/complement(3433. .5049)
/gene="groEL"
CDS
/complement(3433. .5049)
/gene="groEL"
[Genomic island nu Sa 4mw]
/codon_start=1
/transl_table=11
/product="GroEL protein"
/protein_id="BAB95818.1"
/db_xref="GI:21205124"
/transl_table="MVKQLKFSEDARQMLGVDQLANAVKVTIGPKGNVVLKKEFT
APLNTDGVITAKIELEDPYENMGAKLVQVANKNEIAGDGTITATVLAQMIQEG
LKNVTGANPVGLQGDIAVKVAVEALHENSQVKNKEIAQGAISAADEEIGRYI
SEAMEKVGNDGVITIEESNGINTELEVVEGQMDRGQSPYMTVDSQDVALEIRPYI
LVTDKISSFODILPILLEGVQVQSNRPILIVADEVEGDALTIVLNRMGTTAVAVKA
PQGDGRKAMLEDLAILGAOVITDGLDLQSLKASIDMLGTASKVEVTKNTTVVDGD
GDENSIDARVQLKSQIBETSDPDREXLOERLAKLAGGAVIKVGAASETELKRL
RIEDALNSTRAAVEEGIVAGGTALVNVYQVSELEAGDIEGTGVNIVLKALTPVRQ
IAENAGLEGSVIVERLKNAPQNGVFNAAATNENWNNLEAGIVDPTKVTKRSALQHAASVA
AMFLTEAVVASIPEKNAPQNGMGPQGM"
/complement(5125. .5409)
/gene="groES"
CDS
/complement(5125. .5409)
/gene="groES"
[Genomic island nu Sa 4mw]
/codon_start=1
/transl_table=11
/product="GroES protein"
/protein_id="BAB95819.1"
/db_xref="GI:21205125"
/transl_table="MLKIGNRVILIEKKEQETTKSGIVLTDSAKEKNEGIVAVGT
GSLLDNGKTVTEVKEGDRVVFQYAGVEKRDNIYVLNDEEDILAVIE"
/5584. .6327
/gene="MW1955"
CDS
/5584. .6327
/gene="MW1955"
[Genomic island nu Sa 4mw]
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="BAB95820.1"
/db_xref="GI:21205126"
/transl_table="WTRLWASLLTVIIVLSOLFPLIIVKLLPFVQYSGIELTKAVIY
IQVLFLIAATIIILNKLKNPTKLELVKPKVLIIPALLIGPALVMIVQWVSIY
LTOYGGQVSPNTEKLIITARKKIFIEFVSIIGPLLEEVFRKVIIFGELFNAIKGN
RIVAFIATVSVSLIPALAHNDKFIPIVYFGVIFSLAYVWTKRLAVFIHMLQNG
FVVFQILNPEALKKATEQANFIYHIFIP"
/complement(6352. .7611)
/gene="MW1956"
CDS
/complement(6352. .7611)
/gene="MW1956"
[Genomic island nu Sa 4mw]
/codon_start=1
/transl_table=11
/product="hypothetical protein, similar to SdrH"
/protein_id="BAB95821.1"
/db_xref="GI:21205127"
/transl_table="MSYHFKKMLLSTSLILSSSLGLATHVEAKDNLNGEKPPTTN
LNSHNTSPSNSEMMNNEGTTPHESNOTGNEGTSRNDANPDNNKVPDNNQNPST
DKSPDNNQNPSPNPKPDNPKPDPPDKPDPPDKPDPPDKPDPPDKPDPPDKPDPP
DKPKPDPPDKPDPPDKPDPPDKPDPPDKPDPPDKPDPPDKPDPPDKPDPPDKPDPP
ASPSGQOQOQNGNQNQNPNDGDFVQRFALANGAYKYNFYILNKLKGDYD
EVTDEIYVIRKQNSGNAYLNGLOQSNYFRFVFNPLKSRYYRNLDEQVLALIT
GEIGSMFDLKKPEDKPDSPKORSEPEHKDDFTVVVKQEDNKKKSASTAYSKSWLAIVCS
MMVFSIMLFLVFRNKKKNKNSQRR"
/7808. .8434
/gene="MW1957"
CDS
/7808. .8434
/gene="MW1957"

```

```

/Note="ORFID:MW1957"
/codon_start=1
/transl_table=1
/product="conserved hypothetical protein"
/protein_id="BAB95822.1"
/db_xref="GI:21205128"
/translation="MGLPKKDKAMTFDNAMEERRSIYNLKDSSISDDELSMIAHA
VKHVPSSNSQSTRVLLNNNNKFWNTYAILKEVNGENRDEPEPEQKIDNPKHSY
STILYEQDVQVSGLOEOMPNYDFNFAISQTNMHPFAIWTALATKNGISGLQHYN
FLVDEMTSENFNIPKSWKLIQAQMPFDIREAGSEKTFNVEDRFVIKK"
8795..9580
/gene="MW1958"
CDS
8795..9580
/gene="MW1958"
/Note="ORFID:MW1958"

Query Match      87.3%; Score 826.4; DB 1; Length 304050;
Best Local Similarity 98.5%; Pred. No. 1.8e-109;
Matches 855; Conservative 0; Mismatches 11; Indels 2; Gaps 2;

QY 80 ATTACCGAATTTTATACCTTATTTCTTTAGATGAACCTTTTATAACATAGTTGGATAGT 139
Db 262253 ATATGTGATTTTATATCTTATTTGTTTGAATGAACTTTTATACATAGTTGGATAG- 262195
QY 140 TTTCGATTTATACATTAATTAATGTAACCTTCTCTACAACAAGATGTGCATCAGAGAGTG 199
Db 262194 TTTCGATTTTAAATTAATTAATGTAACCTTCTCTACAACAAGATGTGCATCAG-AGGAGTG 262136
QY 200 GTTAAATATGATGATTAATTAATGATTAATGATTTAGTCAACGCAACATTTCAAGTTA 259
Db 262135 GTTAAATATGATGATTAATTAATGATTAATGATTTAGTCAACGCAACATTTCAAGTTA 262076
QY 260 AGAAGTTTTTCAGAGATACAAAAGAGAGTTCAATTTGAACTAAGAAATTTATATTT 319
Db 262075 AGAAGTTTTTCAGAGATACAAAAGAGAGTTCAATTTGAACTAAGAAATTTATATTT 262016
QY 320 TAATCATATTTTGAAGAGTGAGTCTACAGAAATCTCATCTAAAGAGATTCGTAAGTGT 379
Db 262015 TAATCATATTTTGAAGAGTGAGTCTACAGAAATCTCATCTAAAGAGATTCGTAAGTGT 261956
QY 380 CAGAGTTCAAAACCTTACTTATTTAACTAAAGCTTTTACAAAAGCTTAAAGATTTTAAATTTG 439
Db 261956 CAGAGTTCAAAACCTTACTTATTTAACTAAAGCTTTTACAAAAGCTTAAAGATTTTAAATTTG 261896
QY 440 TATCAAGAAAAGAGTTTACAGCAAGCAAGCAAGTATTTGTTTATGTTACAGATACAC 499
Db 261895 TATCAAGAAAAGAGTTTACAGCAAGCAAGCAAGTATTTGTTTATGTTACAGATACAC 261836
QY 500 AAAAGGCAATATTCAAAACCTGATTTTCAAGATTTAGAGATACATTAAGAAATTAATCA 559
Db 261835 AAAAGGCAATATTCAAAACCTGATTTTCAAGATTTAGAGATACATTAAGAAATTAATCA 261776
QY 560 AGGTAATTCGGTTTAAATAACATTTGAACGATAACAATTTTAAATACGAAGTTTATTTAT 619
Db 261775 AGGTAATTCGGTTTAAATAACATTTGAACGATAACAATTTTAAATACGAAGTTTATTTAT 261716
QY 620 CAGCATTTGGGACATAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAGTT 679
Db 261715 CAGCATTTGGGACATAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAGTT 261656
QY 680 AAAGACGACCATACCTTATTTACGCTTATTAAGCTTTGACACAGGTACACTAGTCTTTT 739
Db 261655 AAAGACGACCATACCTTATTTACGCTTATTAAGCTTTGACACAGGTACACTAGTCTTTT 261596
QY 740 TATTTTAATTTTCTTAGAATATCAAGTTTACATCAATTAATTTTCTGCGATATAC 799
Db 261595 TATTTTAATTTTCTTAGAATATCAAGTTTACATCAATTAATTTTCTGCGATATAC 261536
QY 800 TTTCGATGTTTCAAGTATTTTCTCTATAATTTTGTGCGATAGCAAGCAAAATTTCTAACTG 859
Db 261535 TTTCGATGTTTCAAGTATTTTCTCTATAATTTTGTGCGATAGCAAGCAAAATTTCTAACTG 261476
QY 860 CAAACCATGTAGGCAATTGAAATATGACCAACAGCAATCGGATGTATTGCTATAAG 919

```

```

Db 261475 CAAACCATGTAGGCAATTGAAATATGACCAACAGCAATCGGATGTATTGCTATAAG 261416
QY 920 AACTAATGTTTTCATAAACTGAATCGAT 947
Db 261415 AACTAATGTTTTCATAAACTGAATCGAT 261388

RESULT 3
AP0031136/c 307750 bp DNA linear BCT 11-JAN-2003
Staphylococcus aureus subsp. aureus N315 genomic DNA, complete
genome, section 8/10.
AP003136 BA000018
AP003136.2 GI:14349228

Staphylococcus aureus subsp. aureus N315
Staphylococcus aureus subsp. aureus N315
Bacteria; Firmicutes; Bacillales; Staphylococcus.

REFERENCE
1
AUTHORS
Kuroda,M., Ohta,T., Uchiyama,I., Baba,T., Yuzawa,H., Kobayashi,I.,
Cui,L., Oguchi,A., Aoki,K., Nagai,Y., Lian,J., Ito,T., Kanamori,M.,
Matsumaru,H., Maruyama,A., Murakami,H., Hoshino,A.,
Mizutani-Ui,Y., Takahashi,N.K., Sawano,T., Inoue,K., Kaito,C.,
Sekimizu,K., Hirakawa,H., Kuhara,S., Goto,S., Yabuzaki,J.,
Kanahisa,M., Yamashita,A., Oshima,K., Furuya,K., Yoshino,C.,
Shiba,T., Hattori,M., Ogasawara,N., Hayashi,H. and Hiramatsu,K.
Whole genome sequencing of methicillin-resistant Staphylococcus
aureus
Lancet 357 (9264), 1225-1240 (2001)
JOURNAL
MEDLINE
21311952
PUBMED
11418146
REFERENCE
2 (bases 1 to 307750)
AUTHORS
Director-General, Biotechnology Center, Aoki,K., Oguchi,A.,
Hosoyama,A., Nagai,Y., Kuroda,M., Hiramatsu,K. and Kikuchi,H.
Direct Submission
Submitted (30-JAN-2001) Director-General, Biotechnology Center,
National Institute of Technology and Evaluation, Biotechnology
Center; 2Chome 4-10 Nishihara, Shibuya-ku, Tokyo 151-0066, Japan
(E-mail:bio@nig.go.jp, URL:http://www.bio.nig.go.jp/,
Tel:81-3-3481-1933, Fax:81-3-3481-8424)
On Jun 12, 2001 this sequence version replaced gi:13701842.

FEATURES
Location/Qualifiers
1..307750
/organism="Staphylococcus aureus subsp. aureus N315"
/mol_type="genomic DNA"
/strain="N315"
/sub_species="aureus N315"
/db_xref="taxon:158879"
297..1907
/gene="SA1853"
297..1907
/gene="SA1853"
/Note="ORFID:SA1853"
hypothetical protein, similar to DNA mismatch repair
protein MutS"
/codon_start=1
/transl_table=11
/protein_id="BAB43135.1"
/db_xref="GI:13701843"
/translation="MSTNQTFLFVIAIILLTSVIGIVGYSRQRLKSMETLWOTI
SPLEAFIRPNSHDPEYKLYKEFESHLVDDKTWDLNNNAIFKXNNYNTAIGENK
LYACLRGLMSTIKSLSLDFNSQILLFNLSCLFNILSATLKRTEYDDLLKSFYASNVLK
FLCFLPVSFAVIFINSQVILLFNLSCLFNILSATLKRTEYDDLLKSFYASNVLK
QGYTISKIKHAPQEVNFQRTARHLTSVAEVDNEDIGAMVILKLFMLDYLF
HSIQKSTYTHMELKNCDFYAEIDNHYALAMVRLTECYTEPIDNDGIVDFSELT
HPLIADAVANDFSLSONILITGNSAGSTFMAKSIANIILASAIOTVTAKRYVOPG
IVFTSMANADVLGSDGSFMAELKSIKRIVEIPDNOKIYCFIDEIFKGTNTTETRIAS
EVLSPFLHESNRFVIAETHDILAEILLKQRYENHFVNIENNINHFYKIRPGKAN
TRNAIELLKTSPAKTYSERAKDNVNG"
complement(2222..3247)
/gene="SA1854"
complement(2222..3247)

```



```

/gene="SA1854"
/notes="ORFID:SA1854
hypothetical protein, similar to O-sialoglycoprotein
endopeptidase"
/codon_start=1
/transl_table=11
/protein_id="BAB43136.1"
/db_xref="GI:13701844"
/translation="MTKDILILAVETSCDETSYVINGRGRDILSNTVLSQIESKRFQ
GVPEVASRHHVIGITITINEALVADVSDVEDIDAVTEGPGGLIGALLIGNAAKAL
AFAYDKELIPVHHIAGHYANHIEBPLTFPLIALIIVSGGHTELVYKXKHLSPVIGET
RDDAVGAYDKVARTIGLINYPGGQVDRLAEBGEDTYSFPRVMDKSDYDFSGKLS
AVINQLNQRKPIITTEANVYTFQNSVVEVLTTFKAIQACKEYSVORLIVAGGVASN
GLRQSLADQCKNDIQLTIPSPKLTCDNAAMIGVAGHSLYQQGRFADLALNGHSNID
LEYSAAE"
/complement(3240..3704)
/gene="SA1855"
/complement(3240..3704)
/gene="SA1855"
/notes="ORFID:SA1855
hypothetical protein, similar to ribosomal-protein-alanine
N-acetyltransferase"
/codon_start=1
/transl_table=11
/protein_id="BAB43137.1"
/db_xref="GI:13701845"
/translation="MDQSKKEQLNIREMTKEDVQVDFLERRSNDSSWTIDAFYHEI
EQNNFAKYFEFQIIGYLGWLVIDQAITVAIDDNYRGYGLQMLLYKGINYA
SHTCDVMSLEVNVANKVAQHYVNLFGYGGKRNYYGEGEDAMVMVNLND"
/complement(3677..4366)
/gene="SA1856"
/complement(3677..4366)
/gene="SA1856"
/notes="ORFID:SA1856
hypothetical protein, similar to glycoprotein
endopeptidase"
/codon_start=1
/transl_table=11
/protein_id="BAB43138.1"
/db_xref="GI:13701846"
/translation="MNKLRRQLMSNLIDTSNPLSVALMNQNDVLAETTSQKHQ
SVQLMPALSQLPEQSKIAKQOLDALIVAEQGSYTGRIQVTVAKTLYALDVKLYGV
SSLKALATIDHTDKLIVPFDARRQAVYGTIFQWQQLVTLIEDQYMSIALRAFL
EDLNQPFYIGKDIKLQDELQGEVIAQLPNASVMYQLIDKPTDVHTFTPKYHLAEA
ERNWINSQANN"
/complement(4320..4781)
/gene="SA1857"
/complement(4320..4781)
/gene="SA1857"
/notes="ORFID:SA1857"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="BAB43139.1"
/db_xref="GI:13701847"
/translation="MIKINLDNMNOFAMFLVEQLKSGDLLINGDLGAGTKTLTQFI
GKALGVRTINSPTNITKSYRGKNLKLHMDCYRLESDSDGLGDFEFQAITVIE
WSQFIKDLIPATHLSINISTISENTRQIELFAQGHYEIQKEAIIHEFAAH"
5292..6980
/gene="ilvD"
5292..6980
/gene="ilvD"
/notes="ORFID:SA1858"
/codon_start=1
/transl_table=11
/product="dihydroxy-acid dehydratase"
/protein_id="BAB43140.1"
/db_xref="GI:13701848"
/translation="NRSDMIKKXGQHQAPARSLHATGALKSPTDMNKPFVAICNSYID
IYVGHVHLRELADIAKEAREGALIPFEPNTIGVDDGIANGHIGNRYSLPSREIIADA
ASTVNAHMDGVYVTPNCDKITPGMILAAMETNPVPAIFCSGGPMKAGLSAHKALTU
SNWFVAVGAFKSGSISKEEFLMEQNAQCEPTCGSCAGMFTANSACLMEVLGLALPYNG
TALAVSDQREMIROAAPKLIVENIKNDLKPDIIVTREAIDDAFALDAMVGGSTNVLH

```

gene

CDS

```

TLAANEAGIDYDLERINAIKRTPYLSKIAPSSSYSHDVHERAGGVPAIINELMKKD
GTLHPDRITVTGKTLRENKEGKIKNFVHPLDAPYDAQGLSLFNFNIAPKGAIVK
VGVGDPSTKITGTGAIKCFNSHDEAVEAIDNRTVRAGVHVIVRYEGPKGPGVPELAP
TSSI VGRGLGKDVALI TDRTSGATRGIAVGHISPEAASGGPIALIEDGETTIDLTN
RTLNVQPEDVLARRRESLTFFKAKVKTGYLARTALVT SANTGGMQVNPENLI"
7008..8777
/gene="ilvB"
7008..8777
/gene="ilvB"
/notes="ORFID:SA1859"
/codon_start=1
/transl_table=11
/product="acetolactate synthase large subunit"
/protein_id="BAB43141.1"
/db_xref="GI:13701849"
/translation="MSKTQHEVNQINIDPLKMAESLELEQLNEKTLNDRSSSEVLVEA
LLKENVDYLFYGGVAGVLPDYDTFYDGKIKHILARHEQCAVHAAEGYARVSGKTGVV
VTRSGPATNMVTGITDAHCDSLPVFTGVATPGIKGDAFOEADILSMTSPITKQNY
QVKRVEDIPKI VHEAFHVANSRGRKGPVVIDPKDMGLVATNVDLCDEINIPGYEVVTE
PENKIDIDTFISLLEAKKPVVLGAGINQSKNQLLTQFVNKHQIPTVTLLGLGAVP
YEDTLFNGMGHGSYASNMALTECDLLINLGSREDDSLASKPDADAFNAKIVHVDID
PSEINKVHVDIGLIADCKRFLFECUNDKNVETIERSDWVKHCQNNKQKHPFKLGREDQ
VFCKPQOTIEYIKTNGEAI VITDVQGHQWMAAQFYFPKNHGQVWITSGGLGTMGFGI
PSSIGAKLANPKTVVCFVGGGFQMTNQEXALLPEYGLDVKIVLNNGTIGMWKQWQ
DKTFNCRFSHSVFNGQDFPMKMAEAYGVKFLIDKPEQLEEQLEQDAAFAYQGPVLEVR
ISPTAEVTPMVPSPGKSNHEMGL"
8777..9031
/gene="SA1860"
8777..9031
/gene="SA1860"
/notes="ORFID:SA1860
hypothetical protein, similar to acetolactate synthase
small subunit"
/codon_start=1
/transl_table=11
/protein_id="BAB43142.1"
/db_xref="GI:13701850"
/translation="WTRILKQVADQVSTLNRTSAFVRLQYNIDTLVHTHSEQPGIS
NMEIQVDIODDTSLHILIKKLKQINVLTVCEQVLDVNEA"
9168..10172
/gene="ilvC"
9168..10172
/gene="ilvC"
/notes="ORFID:SA1861"
/codon_start=1
/transl_table=11
/product="alpha-keto-beta-hydroxylacil reductoisomerase"
/protein_id="BAB43143.1"
/db_xref="GI:13701851"
/translation="MTTVYDQDVKTDAIOGKIIAVYGSQCHAHQNLKONGYDVV
IGIRPGRSFDKAKEDGDFVPAEAVKQADVIMVLLPDEIQGDVYKNEIEPNLEKHA
LAFAGHFNHFQVIOPPADVDFVLPAPKPGHVRFTFVEGSAVPSLFIQDASQA
RNIALSVAKIGATRAVETITTFKEETETDLFGEQAVLCGGVSKLIQSGFETLVEAGY
QPELAYFEVLHMKLIVDLMEYEGGMENVRYSISNTAEFGDYVSGPRVITPDVKENMKA
VLTDIQNGHNSNFIEDNKNKGFYKLBREHQGHIEKVGRELREMPFIKSKSIEK"
10202..11731
/gene="leuA"
10202..11731
/gene="leuA"
/notes="ORFID:SA1862"
/codon_start=1
/transl_table=11
/product="2-isopropylmalate synthase"

```

Query Match

87.3%; Score 826.4; DB 1; Length 307750;

Best Local Similarity

98.5%; Pred. No. 1.8e-109;

Matches 855; Conservative

0; Mismatches 11; Indels 2; Gaps 2;

CY

80 ATTACCAATTTTATCTATTGTTAGAACTTATACATAGTTGATAGACAT

DB

256270 AATATTGAATTTTATCTATTGTTAGAACTTATACATAGTTGATAGACAT- 256212

140 TTTCGATTAAATACATTAATGTGAACCTTGCTCAACAAGATGTCATCGAAGGAGTG 199  
 Db |  
 256211 TTTCGATTAAATACATGATGAACCTTGCTCAACAAGATGTCATCGAAGGAGTG 256153  
 Qy |  
 200 GTTTAATAATGAGTAAATTAATGATATTAATGATTAGTCAACGCAACATTTCAAGTTA 259  
 Db |  
 256152 GTTTAATAATGAGTAAATTAATGATATTAATGATTAGTCAACGCAACATTTCAAGTTA 256093  
 Qy |  
 260 AGAAGTTTTTCAGAGATACAAAAGAGAGTTCAAATTTGAACTATGCAAGAAATTTATATTT 319  
 Db |  
 256092 AGAAGTTTTTCAGAGATACAAAAGAGAGTTCAAATTTGAACTATGCAAGAAATTTATATTT 256033  
 Qy |  
 320 TAAATCATATTTTAAAGAGTGAAGTCTAAAGAAATCTCATCTAAAGAGATTTGCTAAGTCT 379  
 Db |  
 256032 TAAATCATATTTTAAAGAGTGAAGTCTAAAGAAATCTCATCTAAAGAGATTTGCTAAGTCT 255973  
 Qy |  
 380 CAGAGTTCAACCTTACTATTTAACTAAAGCTTTTCAAAAGCTTAAAGATTTAAATTTGT 439  
 Db |  
 255972 CAGAGTTCAACCTTACTATTTAACTAAAGCTTTTCAAAAGCTTAAAGATTTAAATTTGT 255913  
 Qy |  
 440 TATCAAGAAAAAGAGTTTACAGAGCAAGAAAGAGTATTGTTTATGTTTACAGATACAC 499  
 Db |  
 255912 TATCAAGAAAGAGTTTACAGAGCAAGAAAGAGTATTGTTTATGTTTACAGATACAC 255853  
 Qy |  
 500 AAAAAAGCAATATTCAAAAAGTGAATTCAGAAATTAGAGAAATACATTTAAATTTAAATCA 559  
 Db |  
 255852 AAAAAAGCAATATTCAAAAAGTGAATTCAGAAATTAGAGAAATACATTTAAATTTAAATCA 255793  
 Qy |  
 560 AGGTTAATTTGGTTTAAATCAATTTGAAGCAATTAATTTAATACGAGTATTATTTATT 619  
 Db |  
 255792 AGGTTAATTTGGTTTAAATCAATTTGAAGCAATTAATTTAATACGAGTATTATTTATT 255733  
 Qy |  
 620 CAGCAATGGGCATAAAAATTAACTTAAATTTAAATTTGAAGATGCTTTAAATTTAAAGTT 679  
 Db |  
 255732 CAGCAATGGGCATAAAAATTAACTTAAATTTAAATTTGAAGATGCTTTAAATTTAAAGTT 255673  
 Qy |  
 680 AAGACCAAGCCATACCTTATTTCAGCTTATTAAAGCTTGACACAGGTACACTAGTCTTTT 739  
 Db |  
 255672 AAGACCAAGCCATACCTTATTTCAGCTTATTAAAGCTTGACACAGGTACACTAGTCTTTT 255613  
 Qy |  
 740 TATTTTAAATTTTCTTAAAGAAATCAAGTTTACGATCATAAATATTTTCTCGCATATAGC 799  
 Db |  
 255612 TATTTTAAATTTTCTTAAAGAAATCAAGTTTACGATCATAAATATTTTCTCGCATATAGC 255553  
 Qy |  
 800 TTGATGCTTCCAGATTTTCTTAAATTTTCTTAAATTTGTCGCAATAGCAAAAATTTCTAAGT 859  
 Db |  
 255552 TTGATGCTTCCAGATTTTCTTAAATTTTCTTAAATTTGTCGCAATAGCAAAAATTTCTAAGT 255493  
 Qy |  
 860 CAAAACCATGTGTAGGCAATTTGAGAAATAGCAACAGCAATCGGATGTTTCTTAAAGT 919  
 Db |  
 255492 CAAAACCATGTGTAGGCAATTTGAGAAATAGCAACAGCAATCGGATGTTTCTTAAAGT 255433  
 Qy |  
 920 AACTAATGTTTCAATAAAGTGAATCGAT 947  
 Db |  
 255432 AACTAATGTTTCAATAAAGTGAATCGAT 255405

RESULT 4  
 AP003364/c 348650 bp DNA linear BCT 07-FEB-2002  
 LOCUS  
 DEFINITION Staphylococcus aureus subsp. aureus Mu50 genomic DNA, complete  
 sequence, section 7/9.  
 ACCESSION AP003364 BA000017  
 VERSION AP003364.2 GI:14247707  
 KEYWORDS  
 SOURCE Staphylococcus aureus subsp. aureus Mu50  
 ORGANISM Staphylococcus aureus subsp. aureus Mu50  
 Bacteria; Firmicutes; Bacillales; Staphylococcus.  
 REFERENCE 1  
 AUTHORS Kuroda, M., Ohta, T., Uchiyama, I., Baba, T., Yuzawa, H., Kobayashi, I.,  
 Cui, L., Oguchi, A., Aoki, K., Nagai, Y., Lian, J., Ito, T., Kanamori, M.,  
 Matsumaru, H., Maruyama, A., Murakami, H., Hosoyama, A.,  
 Mizutani, U. Y., Takahashi, N. K., Sawano, T., Inoue, R., Kaito, C.,

Sekimizu, K., Hirakawa, H., Kuhara, S., Goto, S., Yabuzaki, J.,  
 Kanehisa, M., Yamashita, A., Oshima, K., Furuya, K., Yoshino, C.,  
 Shiba, T., Hattori, M., Ogasawara, N., Hayashi, H. and Hiramatsu, K.  
 Whole genome sequencing of methicillin-resistant Staphylococcus  
 aureus  
 Lancet 357 (9264), 1225-1240 (2001)  
 JOURNAL  
 MEDLINE 21311952  
 PUBMED 11418146  
 REFERENCE 2 (bases 1 to 348650)  
 AUTHORS Ohta, T.  
 TITLE Direct Submission  
 JOURNAL Submitted (28-FEB-2001) Toshiko Ohta, University of Tsukuba College  
 of Medical Technology and Nursing, Department of Medical  
 Technology, 1-1-1 Ten-nodai, Tsukuba, Ibaraki 305-8577, Japan  
 E-mail: tohtasakura.cc.tsukuba.ac.jp. Tel: 81-298-53-3454,  
 Fax: 81-298-53-3454  
 COMMENT On May 29, 2001 this sequence version replaced gi:13873637.  
 FEATURES  
 1. 348650  
 Location/Qualifiers  
 /organism="Staphylococcus aureus subsp. aureus Mu50"  
 /mol\_type="Genomic DNA"  
 /strain="Mu50"  
 /sub\_species="aureus Mu50"  
 /db\_xref="taxon:158878"  
 complement(59..235)  
 /gene="SAV1935"  
 complement(59..235)  
 /gene="SAV1935"  
 /codon\_start=1  
 /transl\_table=1  
 /product="hypothetical protein"  
 /protein\_id="BAB58097.1"  
 /db\_xref="GI:14247708"  
 translation="MKHLTKIFVVLAILFLIIGYLOATNHSQGIKLLLAAMFMC  
 AFISRSNDRKNSK"  
 775..2061  
 /gene="SAV1936"  
 775..2061  
 /gene="SAV1936"  
 /note="similar to aspartate transaminase protein"  
 /codon\_start=1  
 /transl\_table=1  
 /product="hypothetical protein"  
 /protein\_id="BAB58098.1"  
 /db\_xref="GI:14247709"  
 translation="MNPALQSLNEQLQSNATFTMLSDLCNNMFYKGLISQSAAEK  
 STTNATIGMATNKGKMFASLSDAMENDLTPDIPYAPQGEIEELDLWOQKWLKD  
 NPEISLDNMSPIVTNALTHGLSLVGLDFVNGDTILLPEHNWNYKLIVNTRNGAIL  
 QTYPLFDKGHTYDLSLEALQCSYKQKVMILNYPNNPTGYTTHKEVTIVDAIKA  
 LADKGTAVVDDAYYGLFYEDVYTSQSLFTALSNLSNAILPVRLDGATKEFFAWGF  
 RVGFTFTGSDQTTKEVLEAKVGLIRSNSSGPLETQSAVHVKNNKQFDKEIQN  
 IQTLKERYEVTKVYADQVHSHWQAYDFNSGYFMAIKVHDVDPPEELRKHLEKYSIG  
 VIALNADIRAFSCVEKEDIPIHVFDSIAKADDLR"  
 complement(2340..4075)  
 /gene="truncated-mapw"  
 complement(2340..2660)  
 /gene="truncated-mapw"  
 /note="SAV1937"  
 /codon\_start=1  
 /transl\_table=1  
 /product="truncated map-w protein"  
 /protein\_id="BAB58099.1"  
 /db\_xref="GI:14247710"  
 translation="MKKKINKNSNNVKFPVTINKFENVSNEFVFNASKTITINDLSIK  
 LKSAVANDQGITKEDIGLAEPRVYKVFKNSSKYVDLKTKEYKDERVFKATDIKKVDI  
 ELKS"  
 complement(2645..4075)  
 /gene="truncated-mapw"  
 /note="SAV1938"  
 /codon\_start=1  
 /transl\_table=1  
 /product="truncated map-w protein"  
 /protein\_id="BAB58100.1"

```

/db_xref="GI:14247711"
/translation="MKFKSLITTTTLAGVIASTGANFNTNEASAAKPLDKSSSTLHH
GHSNIQIPTYITVNGTSQNILSSTFNKQNIQISYKDIENKVSVLVFNRGISDIDLRL
SKQAEYTFHFKNGKRVLDKSGIYVADLINTSDIKAI SVNVDTKQPKOKAKANQV
PYITVNGTSQNILSSTFNKQNIQISYKDIENKVSVLVFNRGITDVLRLSKQAKYT
VFNKGTAKVLDKSGIYVADLINTSDIKAI SVNVDTKQPKOKAKANQVQVPSINLN
GTSNINLSNFSKPNWYNKQNLQISQISVLKHDRGISEQDLKYAKAYIVYFPGNG
KRILQKNVTANLVHARDVKRIEITVTGTAKADRIYVYITAVNGTSTPILSRUK
ISNQLISYKYLNDKVKSVLKSERGISDLDLKFAKQAKYTVYFNGKKQVNVNLSKDP
TPNLSAKDIXKIDIVQYTKSKK"
4418..4660
/gene="truncated-hlb"
CDS
4418..4660
/gene="truncated-hlb"
/translation="truncated 6449-6450 beta-hemolysin"
/db_xref="GI:14247712"
/translation="MLHLPCKLYKRSNDGKTKNSLKKVATLALANLLLVGALTD
NSAKAESKDDTDLKLVSHNVMJSTVLVFNWRLIT"
5032..5211
/gene="SAV1940"
5032..5211
/gene="SAV1940"
/translation="Bacteriophage phiN315"
/db_xref="GI:14247713"
/translation="MKFKTCKSNNTFKRSQENNSSEKGLWGIMLLFLLI
LFGITACSSSHFIN"
5370..5543
/gene="SAV1941"
5370..5543
/gene="SAV1941"
/translation="hypotheical protein"
/db_xref="GI:14247714"
/translation="MGIFVAILYGMKRAYWLNLLIFLTNIFYFLHFCVLLLLKYIG
KLFIVEGCVLFT"
5834..6184
/gene="SAV1942"
5834..6184
/gene="SAV1942"
/translation="Bacteriophage phiN315"
/db_xref="GI:14247715"
/translation="MKIRSLIAGTLAIVLSPVLNLDKNEAQAISTLSPTNSYQNE
KLANELKSLDLNVELATSLNYYKRTIKISQKAMAYALKKDFKMSSEAKYQLQ
KIYNIDEALKSKY"
6695..7162
/gene="truncated-lytA"
6695..7162
/gene="truncated-lytA"
/translation="truncated 6449-6450 beta-hemolysin"
/db_xref="GI:14247716"
/translation="MLAIDILFALHSIHISQHRNITVATPHLKWGVYFKWCVSGIS
NVKTPAKENKSPVAGTLDKNNYFKKGTGYTVANVKNVNDGYSYTSNRTGVL
PNNATIKVDGACINGYRWITVIANNQRCVIATGEVDKAGNRISSEFNFSAL"
7682..8173
/gene="truncated-lytA"
8173..8173
/gene="truncated-lytA"
/translation="truncated 6449-6450 beta-hemolysin"
/db_xref="GI:14247717"
/translation="MLKRSLLFLVLLLFSSSITNEVSASSSFDKGYKKGDDASY
FEPTGYLMVNTGVGKGNELLSPHYVEFPKPGTTLTKKIEYVVEWALDATAKKE
FRVYELDPKAKIEVTVYDKNKKBEKSPITEKGFVFDLSEHINKNPGNLTITKVI
EKK"
8388..9119
/gene="SAV1945"
8388..9119
/gene="SAV1945"
/translation="Bacteriophage phiN315"
/db_xref="GI:14247718"
/translation="MKTYSEARLEWYQGRYIDFDGWYQYCADLAVDIYWLLEIR
MMGNKADAINDPKXWATVVENTPSEVPOIGDVAFTKGIYQYGHIGLVFNGNTNQ
FLILEONYDGNANTPAKLWDNVYGGCTHIFRKYKSEGLMKNITNKVKPAPKAVGKS
ASKITVSGKAPYNLKSAGYFNKIDGIGALSATRYGDNRTNRYRFDVQAVVAPGTL
IYVFIIDGWCRIYWNHNWICYE"
9131..9385
/gene="SAV1946"
9131..9385
/gene="SAV1946"
/translation="Bacteriophage phiN315"
/db_xref="GI:14247719"
/translation="MINWKIRMKQSFVALLSAIFLFAQNTAKAIGYDIQVYVEQLT

```

```

CDS
/gene="sak"
/complement(7682..8173)
/gene="sak"
/notes="Bacteriophage phiN315
SAV1944"
/codon_start=1
/transl_table=11
/product="staphylokinase precursor"
/protein_id="BAB58106.1"
/db_xref="GI:14247717"
/translation="MLKRSLLFLVLLLFSSSITNEVSASSSFDKGYKKGDDASY
FEPTGYLMVNTGVGKGNELLSPHYVEFPKPGTTLTKKIEYVVEWALDATAKKE
FRVYELDPKAKIEVTVYDKNKKBEKSPITEKGFVFDLSEHINKNPGNLTITKVI
EKK"
8388..9119
/gene="SAV1945"
8388..9119
/gene="SAV1945"
/translation="Bacteriophage phiN315"
/db_xref="GI:14247718"
/translation="MKTYSEARLEWYQGRYIDFDGWYQYCADLAVDIYWLLEIR
MMGNKADAINDPKXWATVVENTPSEVPOIGDVAFTKGIYQYGHIGLVFNGNTNQ
FLILEONYDGNANTPAKLWDNVYGGCTHIFRKYKSEGLMKNITNKVKPAPKAVGKS
ASKITVSGKAPYNLKSAGYFNKIDGIGALSATRYGDNRTNRYRFDVQAVVAPGTL
IYVFIIDGWCRIYWNHNWICYE"
9131..9385
/gene="SAV1946"
9131..9385
/gene="SAV1946"
/translation="Bacteriophage phiN315"
/db_xref="GI:14247719"
/translation="MINWKIRMKQSFVALLSAIFLFAQNTAKAIGYDIQVYVEQLT

```

```

Query Match 87.1%; Score 824.8; DB 1; Length 348650;
Best Local Similarity 98.4%; Pred. No. 2.9e-109;
Matches 854; Conservative 0; Mismatches 12; Indels 2; Gaps 2;
Qy Db 80 ATTACCGAATTTTATACCTTATTTTGTAGAACTTATTAACATAGTTGATAGAGT 139
|||
|||
|||
339027 AATATTGAATTTTATACCTTATTTTGTAGAACTTATTAACATAGTTGATAGAG- 338969
Qy 140 TTTCGATTTAATACATTAATGTGACCTTGTACACCAAGATGTGATCAGAGAGTG 199
|||
|||
|||
338968 TTTCGATTTAATAAATTACATGTGACCTTGTACACCAAGATGTGATCAG-AGGAGTG 338910
Qy 200 GTTTAATAAGTAAATTAATGAATTAATGAATTAATGAATTAATGAATTAATGAATTA 259
|||
|||
|||
338909 GTTTAATAAGTAAATTAATGAATTAATGAATTAATGAATTAATGAATTAATGAATTA 338850
Qy 260 AGAAGTGTTCAGAGATACAAAAAAGAGTCAATTTGAACACTATGAAGAAATTTATATTT 319
|||
|||
|||
338849 AGAAGTGTTCAGAGATACAAAAAAGAGTCAATTTGAACACTATGAAGAAATTTATATTT 338790
Qy 320 TAAATCATATTTAAGAGTGTGCTTAACCAATCTCATCTAAGAGATGTGTAAGTGCT 379
|||
|||
|||
338789 TAAATCATATTTAAGAGTGTGCTTAACCAATCTCATCTAAGAGATGTGTAAGTGCT 338730
Qy 380 CAGAGTTCAAAACCTTACTATTATTAACCTTACAAAAGCTTACAAAAGCTTAAAAATTGT 439
|||
|||
|||
338729 CAGAGTTCAAAACCTTACTATTATTAACCTTACAAAAGCTTACAAAAGCTTAAAAATTGT 338670
Qy 440 TATCAAAAGAAAGAGTTTACAGACGAAAGACAGTATTGTTTATGTTTACAGATACAC 499
|||
|||
|||
338669 TATCAAAAGAAAGAGTTTACAGACGAAAGACAGTATTGTTTATGTTTACAGATACAC 338610
|||
|||
|||

```

```
QY 500 AAAAAACAAATATTCAAAAACATGATTCAGAAATTAAGAAATACATTAATAATTAATCA 559
Db 338609 AAAAAACAAATATTCAAAAACATGATTCAGAAATTAAGAAATACATTAATAATTAATCA 338550
QY 560 AGTTAAATTCGTTTAAATTAACATTAAGCATTAACAAATTAATTAACGAGTTATTTAT 619
Db 338549 AGTTAAATTCGTTTAAATTAACATTAAGCATTAACAAATTAATTAACGAGTTATTTAT 338490
QY 620 CAGCATTTGGGACATAAAAATTAACCTTAAATTTAAATTTAAATTTAAATTTAAATTTAAATTT 679
Db 338489 CAGCATTTGGGACATAAAAATTAACCTTAAATTTAAATTTAAATTTAAATTTAAATTTAAATTT 338430
QY 680 AAAGACCCACCATACCTTATTTTACGCTTATTAAGCTTGACACAGGTACACATGCTTTT 739
Db 338429 AAAGACCTACCCATACCTTATTTTACGCTTATTAAGCTTGACACAGGTACACATGCTTTT 338370
QY 740 TATTTTAAATTTTCTTAGAAATCAAGTTTACGATCAATAATATTTTCTCGATATAGC 799
Db 338369 TATTTTAAATTTTCTTAGAAATCAAGTTTACGATCAATAATATTTTCTCGATATAGC 338310
QY 800 TTGGATGTTTCAAGTATTTTCTCTATTAATTTGTGCGATAGCAAAAAATTTCTAAGT 859
Db 338309 TTGGATGTTTCAAGTATTTTCTCTATTAATTTGTGCGATAGCAAAAAATTTCTAAGT 338250
QY 860 CAAAACCATGTTAGCAATTCAGAAATAGCAACACGACATCGGATGTTTGTCTATAAG 919
Db 338249 CAAAGCCATGTTAGCAATTCAGAAATAGCAACACGACATCGGATGTTTGTCTATAAG 338190
QY 920 AACTAATGTTTCAATAAATGAATCGAT 947
Db 338189 AACTAATGTTTCAATAAATGAATCGAT 338162

RESULT 5
LOCUS AX620460 384 bp DNA linear PAT 20-FEB-2003
DEFINITION Sequence 3423 from Patent WO02094868.
ACCESSION AX620460
VERSION AX620460.1 GI:28449549
KEYWORDS
SOURCE Staphylococcus aureus
ORGANISM Staphylococcus aureus
Bacteria; Firmicutes; Bacillales; Staphylococcus.
REFERENCE 1
AUTHORS Masignani,V.C., Mora,M.C. and Scarselli,M.C.
TITLE Staphylococcus aureus proteins and nucleic acids
JOURNAL Patent: WO 02094868-A 3423 28-NOV-2002;
Chiron Spa (It)
FEATURES
Location/Qualifiers
1..384
/mol_type="genomic DNA"
/db_xref="taxon:1280"
BASE COUNT 167 a 44 c 58 g 115 t
ORIGIN
Query Match 39.0%; Score 369.8; DB 6; Length 384;
Best Local Similarity 99.2%; Pred. No. 1.6e-43;
Matches 382; Conservative 0; Mismatches 2; Indels 1; Gaps 1;
QY 168 TTGCTACACAGAGTGTGCATCAGAGGAGTGGTTTAAATATGATGATTAATATGATAT 227
Db 1 TTGCTACACAGAGTGTGCATCAG-AGGAGTGGTTTAAATATGATGATTAATATGATAT 59
QY 228 TAATGATTTAGTCAACGCAACATTTCAAGTTTAAAGATTTTTCAGAGATACAAAAAGAA 287
Db 60 TAATGATTTAGTCAACGCAACATTTCAAGTTTAAAGATTTTTCAGAGATACAAAAAGAA 119
QY 288 GTTCAATTTGACATGAGAAATTTATTTTAAATCAATTTTAAAGTGTGATCTAA 347
Db 120 GTTCAATTTGACATGAGAAATTTATTTTAAATCAATTTTAAAGTGTGATCTAA 179
QY 348 CGAAATCTCATCTAAAGAGATTGCTAAGTGTGCAGATTCAGATTCAAACCTTACTATTTAACTAA 407
```

```
Db 180 CGAAATCTCATCTAAAGAGATTGCTAAGTGTGCAGAGTTCAAACCTTACTATTTAACTAA 239
QY 408 AGCTTTACAAAAGCTAAAGATTTAAATTTGTTATCAGAGAAAGAAAGTTTCAAGACGA 467
Db 240 AGCTTTACAAAAGCTAAAGATTTAAATTTGTTATCAGAGAAAGAAAGTTTCAAGACGA 299
QY 468 AAGAACAGTTTATTTGTTTATCAGATACACAAAAGCAAAATATTTCAAAAACCTGATTTTC 527
Db 300 AAGAACAGTTTATTTGTTTATCAGATACACAAAAGCAAAATATTTCAAAAACCTGATTTTC 359
QY 528 AGAATTAGAGATACATTAATAAT 552
Db 360 AGAATTAGAGATGCAATTAATAAT 384

RESULT 6
LOCUS AE016750/c 30698 bp DNA linear BCT 01-JAN-2003
DEFINITION Staphylococcus epidermidis ATCC 12228, section 7 of 9 of the complete genome.
ACCESSION AE016750 AE015929
VERSION AE016750.1 GI:27316220
KEYWORDS
SOURCE Staphylococcus epidermidis ATCC 12228
ORGANISM Staphylococcus epidermidis ATCC 12228
Bacteria; Firmicutes; Bacillales; Staphylococcus.
REFERENCE 1 (bases 1 to 30698)
AUTHORS Zhang,Y., Ren,S., Li,H., Fu,G., Lu,L., Lu,G., Jia,J., Tu,Y., Qin,Z., Chen,Z. and Wen,Y.
TITLE Direct Submission
JOURNAL Submitted (05-NOV-2002) Chinese National Human Genome Center at Shanghai, 250 Bi Bo Road, Shanghai 201203, China
FEATURES
Location/Qualifiers
1..30698
/organism="Staphylococcus epidermidis ATCC 12228"
/mol_type="genomic DNA"
/strain="ATCC 12228"
/db_xref="ATCC:12228"
/db_xref="taxon:116280"
complement(114..186)
/product="rRNA-Lys"
complement(190..264)
/product="rRNA-Gln"
complement(269..352)
/product="rRNA-Tyr"
complement(369..444)
/product="rRNA-Val"
complement(451..522)
/product="rRNA-Glu"
complement(524..598)
/product="rRNA-Asn"
complement(612..726)
/product="5S ribosomal RNA"
complement(886..13713)
/product="23S ribosomal RNA"
complement(2718..2831)
/locus_tag="SEI755"
complement(2718..2831)
/locus_tag="SEI755"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAO05396.1"
/db_xref="GI:27316221"
/transl_table="MMIGRALLFORGAPLGLPNSDKLRMPINLTWESEHC"
complement(3977..5530)
/product="16S ribosomal RNA"
complement(5985..7052)
/locus_tag="SEI756"
complement(5985..7052)
/locus_tag="SEI756"
/codon_start=1
```

```

/transl_table=11
/product="ATP-binding Mrp-like protein"
/protein_id="AA005397.1"
/db_xref="GI:27316222"
/translation="MLITVQVKEITVGALKOPIIDVPLRESEGIYDVSIKDNINHVSVK
VANAQGGQQLDQIAVAKLKEGANTVIGRFEELPSSEVVERIYKSEKPKTIEE
LLSQNFVEFISASGGVGVKSTAVYNLAVALAREKKYGLVDADIYGFSDPMGI
DERGIDGKIIIPVERHGKVISNAFFVEENAPVIMRGMGLKMLTNFFTEVQWGLD
YLLDLDPGGDVALDVHMLPSSKEIIVTTPHTAFAVARAGAMAKHTHTILGVI
ENMSYFSEKGTGKEVYFGKGGKLSDELETQLFALPLEQETWPNENDSPSIIYQSD
DRUGELYILARKVIATQXQ"
complement(7203..8642)
/locus_tag="SE1757"
CDS
complement(7203..8642)
/locus_tag="SE1757"
/codon_start=1
/transl_table=11
/product="multidrug resistance protein"
/protein_id="AA005398.1"
/db_xref="GI:27316223"
/translation="MASYEASKTKRNLIYVMSIAFAVILNOTLNLALPHIMREIN
NENTSOWLIITGFMVNGVMIPLTAYLMDKVKTRPYLYLGAMGFLGISIVAAIAPNG
VLMRTAVTQAGLIMPLMQFTLTFLFSKEHGFAMGLAGLVIQAPALGPTFTGLI
IDNASVPRFVIIVGIALVPEIFGFVSISVNTTKETKLDKRSVIYSLTGFGLMYAF
SSAGNLGFSNPIVCSVLISLIIGIVKQIITISNPLNLKIPFNKIPFESITSMI
IMLSWGPALLIPYQNALGSLLSGLVIMPQAILNGIMSIVTGFYDKYGRPMI
FTGFIILISCTFLFCFLKVDTSYTYLIVAIIRMFVSLMLPINTTGINALKTEDIS
HGTAIMFRGWAGSLGTALMTVMSIGAQWVSSESHASKEMIQRQSVAVDVSPA
LVTVFVIAFVLAIPKEPNHLTHNSRKV"
complement(8912..9382)
/locus_tag="SE1758"
CDS
complement(8912..9382)
/locus_tag="SE1758"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AA005399.1"
/db_xref="GI:27316224"
/translation="WMKFLKNSVHLLVTLVLIVIFVISGAIPLTFLGFLYGLSRIL
IYHLGDFSXNGFYDNIYYGSYVILGYFTLFSIEHLMDFKKNLPKNFPFQGINPH
LISYIVTIMFYFVHYHVHVNHFWMIIIGFLFYCKEVFYPESKNLNNKK"
complement(9570..10907)
/locus_tag="SE1759"
CDS
complement(9570..10907)
/locus_tag="SE1759"
/codon_start=1
/transl_table=11
/product="multidrug transporter"
/protein_id="AA005400.1"
/db_xref="GI:27316225"
/translation="WNLSIITVMALILIMEMAAETSLISLALPTIKNSLNAGNLVS
LVTVFIALVIANPIVGLMSRKFIIYIAVGVLLFALGSLMSGLSQTTFLLISRT
VQGFAGVMSLSQIVPKLAFETIPRYKIMGVSVMGSIISIGLFGALIEPASWH
WLFYNIPIAVIILVMTLPHFDETOVQQRFDIKGLIIFYFIALLMPLGNQHH
IIFNFILALVALLWLFKENSIEQPLTKFNFNISIVLVFITDLIAITLMGYNL
YTPVYLOKLSISQSGSVIPPSVAMITLNFNLAKIEAFTKTYLICSFFVLLVS
SLIMVGLKLPILIAFVAVFAGLSFGYITKDSVIVQETSPTSPKNKMSFYALTKNL
GSSVSTINGVYALNVLFGSNLHNVLGLVLIITAVCLIVMWTLYKSNITIQS"
complement(11146..11829)
/locus_tag="SE1760"
CDS
complement(11146..11829)
/locus_tag="SE1760"
/codon_start=1
/transl_table=11
/product="hemolysin III"
/protein_id="AA005401.1"
/db_xref="GI:27316226"
/translation="MSQSKRKNDSSVETFKDIIPLTFCEIGNAASHCAALTLFI
LFAAVHSFNGGTEISVSVISIFMNFISSTIYHSMQNTSHKYLIRIIDSMI
VVAISGTYFVLLSVGGWGLWLTLLWGLTGLWGLIKYSIATKVNHLRSIIVLVMG
WVGLIIFLPIIINETSWMFIFFIFLOGLSVTIGATYQAQNKPFYHMIWHIIVLASFL
HMLGIFYN"
complement(11848..13035)

```

```

/locus_tag="SE1761"
complement(11848..13035)
/locus_tag="SE1761"
/codon_start=1
/transl_table=11
/product="UDP-N-acetylglucosamine pyrophosphorylase"
/protein_id="AA005402.1"
/db_xref="GI:27316227"
/translation="MLDKNQLEKNQEHLYEYKLMNSNKNALDEKVDQNLAIQID
LYQDLYVNRKTIDDSVSEVYKVSRLNEERHTYQKGYEARNGEFAVLLMAGY
QTRGLYKPGKSGFEIETGSLFELCAROLIRLKEETHTINWYIMTSINHKDTTEYF
KQHKYNDANHIHFKNODNIVALSEEGKLVLRDGHIMETPNGGVFKSLKKACYL
DKQQDHVYIPLNNIDNVLYKLDPLFAGFTVTSKDIOTSKTIQPKSESGLVNV
DCKQIVLEYSELDITDVNFKNANIGIRAFKLGFTISAVDELPHLAIKQLQOLDEN
FGVERPTLKFLFYFDIPRYGTSFVTLQVPREEFSLPKNKGKDSVHTATEDLKRK
DLI"
complement(13090..13599)
/locus_tag="SE1762"
CDS
complement(13090..13599)
/locus_tag="SE1762"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AA005403.1"
/db_xref="GI:27316228"
/translation="MTGKTHASGCGLVGAITTOYFHTDFTSIVLSVSIILPDI
CHQKIGKGRFLRTSFVRILFGHRTFFHSLLEIGISFLFYIOTPMYMYVIVIGM
FSHVLDIUTPRGVKLLYPLPFIIVSPHFHKTGGLVDVSLALASUGAIYTLFQPYLN
TMMHWYLIK"
complement(13776..14036)
/locus_tag="SE1763"
CDS
complement(13776..14036)
/locus_tag="SE1763"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AA005404.1"
/db_xref="GI:27316229"
/translation="MAMTVKNDNEVRIQWRVADIKIPNNEKNVTQDQDIHAPBEEN
GXEISIGTGTGKTNKLVLDTOHLYIITQNDQKYNELTK"
complement(14189..14503)
/locus_tag="SE1764"
CDS
complement(14189..14503)
/locus_tag="SE1764"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AA005405.1"
/db_xref="GI:27316230"
/translation="MFVVTNRITVKKGYAKQMAPNFTKGGPIESLKGFGIEVWQIDK
RR"
complement(14682..16055)
/locus_tag="SE1765"
CDS
complement(14682..16055)
/locus_tag="SE1765"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AA005406.1"
/db_xref="GI:27316231"
/translation="XMGFRKVDIVGRIRKFIITMSLIGILFLVPIVQDGKQOTLP
IAFLAGLKDMLGGINPILVITIVSGILITLCSTYIKNKNLPQGLMSAFNVKIGW
LVLRVLAFFSMLTFLNIGPEIKSDEDTGLFVSLPLTAVLFAAIFLLMLYIG
LLELGLIPRIPMRPLTPGRSTVDNLSAFIGDGTGVLITSRQYGEYISREATV
ISTFVSWSITFAIVIAETIRMODQFFYLVIVISLIAAMIRPIWPKNPIDSYA
KEYSEVARNQLPEGTALKYGDLEATYGVKSPGKFLISGFKVDMWVILPVV
MSGTIATITANTPFIIGKFPVVLHLLQIPEAHSQOTILIGFADMFLPSILIE
GVNDTRFVIGALSISQLVYLSEVGVIGSKIPVSIKLFMIFLIRITLITLIAL
LAHLFTG"
complement(16352..17317)
/locus_tag="SE1766"

```

```

Query Match      32.4%; Score 307; DB 1; Length 300698;
Best Local Similarity 80.6%; Pred. No. 1.6e-35;
Matches 383; Conservative 0; Mismatches 90; Indels 2; Gaps 2;

QY 80 ATTACCGAATTTTATACATTTATTTGTTTGAAGTGAACATTTTATACATAGTTTGGATAGAGT 139
DB 97851 AATATTGAATCTTATACATTTATTTGTTTAAATGTACCTTGTGAACATGGCGGAGAAA-T 97793

QY 140 TTTCCGATTTAATACATTAATGATGCACTTGTCTACCAAGATGTGCATCAGAAGGAGT 199
DB 97792 TTTCCGATTTAATACATTAATGATGCACTTGTCTACCAAGATGTGCATCAGAAGGAGT 97734

QY 200 GTTAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 259
DB 97733 GTTAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 97674

QY 260 AGAAGTTTTCAGAGATACAAAAAGAGTTCAATTTGAACATGATGAAGAAATTTATATTT 319
DB 97673 AAAAAATTTTTCAGAGATACAAAAAGAGTTCAATTTGAACATGATGAAGAAATTTATATTT 97614

QY 320 TAAATCATATTTTAAAGAGTGTCTAAAGAAATCTCATCTAAAGAGATTTCTTAAGTCT 379
DB 97613 TAAATCATATTTTAAAGAGTGTCTAAAGAAATCTCATCTAAAGAGATTTCTTAAGTCT 97554

QY 380 CAGAGTTCAACCTTACTATTTAACTAAAGCTTTTCAAAAAGCTTAAAGATTTAAATTTGT 439
DB 97553 CAGAGTTTAAACCGTATTATTTAACTAAAGCTTTTCAAAAAGCTTAAAGATTTAAATTTGT 97494

QY 440 TATCAAGAAAGAGAGTTTCAAGAGCAAGAGATTTATTTTATTTATTTATTTATTTATTT 499
DB 97493 TATCTTAAAGAGAGAGCTGTATGATGATGATGATGATGATGATGATGATGATGATGAT 97434

QY 500 AAAAAAGCAATATTCAAAAAGCTTTTCAAGATTTAGAGATTAATCAATTTAAAAATTA 554
DB 97433 AACGAGAAAAATAAAAAATTTGATTTTGAAGATTTAGAAAACTATATCAAAATTAATA 97379

RESULT 7
AX620416/c
LOCUS      AX620416      834 bp      DNA      linear      PAT 20-FEB-2003
DEFINITION Sequence 3379 from Patent WO02094868.
ACCESSION  AX620416
VERSION     AX620416.1  GI:28449527
KEYWORDS
SOURCE      Staphylococcus aureus
ORGANISM    Staphylococcus aureus
REFERENCE   1
AUTHORS     Masighani, V.C., Mora, M.C. and Scarselli, M.C.
TITLE       Staphylococcus aureus proteins and nucleic acids
JOURNAL     Patent: WO 02094868-A 3379 28-NOV-2002;
            Chiron Spa (IT)
FEATURES
            Location/Qualifiers
            1..834
            /organism="Staphylococcus aureus"
            /mol_type="genomic DNA"
            /db_xref="taxon:1280"
BASE COUNT   296 a 133 c 126 g 279 t
ORIGIN

Query Match      21.1%; Score 199.4; DB 6; Length 834;
Best Local Similarity 99.5%; Pred. No. 2.2e-19;
Matches 200; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 747 ATATTTCTTAGAATCAAGTTTACGATCATATAATTTTCTCGCATATAGCTTTGGAT 806
DB 834 ATATTTCTTAGAATCAAGTTTACGATCATATAATTTTCTCGCATATAGCTTTGGAT 775

QY 807 GGTTCAGATTTTCTCTATAATTTGTGCGCATAGCAAAAAATCTCACTCAAAACC 866
DB 774 GGTTCAGATTTTCTCTATAATTTGTGCGCATAGCAAAAAATCTCACTCAAAACC 715

QY 867 ATGTGTAGGCAATGAGAAATAGCAACACGATCGATGTTTCTGCTATAGAACTAAT 926

```

```

DB 714 ATGTGTAGGCAATGAGAAATAGCAACACGACAAATGTTATGCTATAGAACTAAT 655
QY 927 GGTTCATAACTGAATCGAT 947
DB 654 GGTTCATAACTGAATCGAT 634

RESULT 8
AX599046/c
LOCUS      AX599046      8056 bp      DNA      linear      PAT 14-FEB-2003
DEFINITION Sequence 386 from Patent WO02077272.
ACCESSION  AX599046
VERSION     AX599046.1  GI:28399186
KEYWORDS
SOURCE      synthetic construct
            synthetic construct
            artificial sequences.
ORGANISM
REFERENCE   1
AUTHORS     Berlin, K., Braun, A., Distler, J., Guebig, D., Howe, A., Mueller, J.,
            Olek, A., Piepenbrock, C., Adorjan, P., Grabs, G., Lesche, R., Leu, E.,
            Lewin, A., Lipscher, E., Maier, S., Model, F., Mueller, V., Otto, T.,
            Pelet, C. and Ziebarth, H.
TITLE       Methods and nucleic acids for the analysis of hematopoietic cell
            proliferative disorders
JOURNAL     Patent: WO 02077272-A 386 03-OCT-2002;
            Epigenomics AG (DE)
FEATURES
            Location/Qualifiers
            1..8056
            /organism="synthetic construct"
            /mol_type="genomic DNA"
            /db_xref="taxon:32630"
            /note="chemically treated genomic DNA (Homo sapiens)"
BASE COUNT   3711 a 0 c 371 g 3974 t
ORIGIN

Query Match      10.2%; Score 96.2; DB 6; Length 8056;
Best Local Similarity 46.7%; Pred. No. 5.3e-05;
Matches 378; Conservative 0; Mismatches 423; Indels 8; Gaps 2;

QY 46 TTTGAAATATAAAAAATATCAATAGTTGGAGTCTATACCGAATTTTATACCTATTATTTGT 105
DB 1493 TTTAATAATTAATTTTAAATATAATTTTATTAATAAATAATTTTATTTTATTTT 1434

QY 106 TTAGAATGAACCTTTTATAACATAGTTGGATAGATTTTCGATTTAATCAATTAATGTGAA 165
DB 1433 AAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATA 1374

QY 166 CCGTCTACACACAGATGTGCATCAGAAGGAGTGGTTTAATAAATAGTAAATTAATGAT 225
DB 1373 ATTTTAAATAAAAAATTAATAATTTAAATTTTATTTTAAATAAATAATTAATAATTTT 1314

QY 226 ATTAATGATTTAGTCAACGCAACATTTCAAGTTTAAGAAGTTTTCAGAGATACAAAAAG 285
DB 1313 TATATTTATTTAATTTTAAATAATATTAATTTTATATTTTATATTAATTAATTAATCA 1254

QY 286 AAGTTCAATTTGAACATATGAAGAAATTTATATTTTAAATCAATTTTAAAGAGTGTCT 345
DB 1253 TTTTATTTTATCAAAAAATTTATTTTATTTAATAAATAATTAATAATTAATAATTAATA 1194

QY 346 AACGAAATCTCATCTAAAGAGATTTGCTAAGTGTCCAGAGTTCAACCTTACTATTTTAACT 405
DB 1193 ATTTTAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1134

QY 406 AAGCTTTACAAAAGCTAAAGATTTTAAATTTGTTATCAAGAAAGAAAGCTTTCAAGAC 465
DB 1133 AAAAAATATAAAAAACCAATATATAACATTTTAAACATTTAAAAAATAAAAAAATAAAAA 1074

QY 466 GAAAGAACAGTTATTTGTTTATGTTTACAGATACACAAAGCAATATTTCAAAACTGATT 525
DB 1073 CAAATTTTATTTTATTTTACATTAATTTTCAAAATTTAAAAATAAATAAATAAATAA 1015

QY 526 TCAGAAATGAGAAATACATTTAAAAAATAAATAATCAAGGTTAATTCGCTTTAATAACATGA 585

```

```

1014 -----TAAATAATTTTAAAAAATATTAATATATTTTATTCATTTAAATAAA 961
Db
586 ACGATAACAATTATTAACAGAGTATTTATTCAGCATGGGACATAAAATTAACCTTA 645
Qy
960 AAAATAAAAAATTTATTTTAAAAAATTTTATTAACAAAAACACAAAAAATTAATAA 901
Db
646 AAAATTAATATTCAGAGTGGTTAAATTAAGTTAAAGACGACGATACCTTATTTTCAGC 705
Qy
900 TAAATTTAAATTAACAAATTTTATTAACAAAAAATAAAAAAACAATTTTITTTT 841
Db
706 TTATTAAGCTTCACACAGGTACTAGTCTTTTATTTATTAATATTTCTTAGAAAAATCA 765
Qy
840 TAAATAAAATATTTTAAATAATCATTTTATAAATTTTATTTTAAATATTAATAATAT-A 782
Db
766 AGTTACAGCATATAATTTTTCGCGATATAGCTTGGATGGTTCACAGTATTTTCTCT 825
Qy
781 ATTTTATTAATTTTAAATTAATTTTAAATTTTAAATTTTAAATTTTAAATTAATAATATTTT 722
Db
826 ATAATTTTGTGCGATAGCAAAAAATCT 854
Qy
721 AAATTAATTTTAAATAATAAAAAAATTTT 693
Db

RESULT 9
AC125567/c
LOCUS
DEFINITION Rattus norvegicus clone CH230-9H22, *** SEQUENCING IN PROGRESS ***,
4 unordered pieces.
ACCESSION AC125567
VERSION AC125567.5 GI:24817949
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 170627)
Muzny,D.Marie, Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
Allen,C., Allen,H., Alibrooks,S., Amin,A., Anguiano,D.,
Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
Delgado,O., Denson,S., Detamo,C., Ding,Y., Dinh,H., Divya,K.,
Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
Gebregorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
Gunaratne,P., Haaland,W., Hamill,C., Hamilton,C., Hamilton,K.,
Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hughes,M.,
Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
Lorensheva,L., Loulseghe,H., Lozado,R.J., Lu,X., Ma,J.,
Maheshwari,M., Mahindaratne,M., Mahmoud,M., Malloy,K., Mangum,A.,
Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
Mawlinsey,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
Nwaokemelame,O., Okwunwo,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
Pascernak,S., Paul,H., Perez,A., Perez,L., Pfankoch,C.,
Pioppfer,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,

```

```

Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
Sanders,W., Savery,G., Scherer,S., Scott,G., Shatman,S., Shen,H.,
Shetty,J., Shvarcbeyn,A., Sisson,I., Sitter,C.D., Smajd,D.,
Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
Steinle,M., Strong,R., Sutton,A., Svatek,A., Taber,P., Taylor,C.,
Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
Valas,R., Vera,V., Villalana,D., Waldron,L., Walker,B., Wang,J.,
Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
Williams,G., Willison,R., Wleczek,R., Wooden,H., Worley,K.,
Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
Weinstock,G. and Gibbs,R.A.
Direct Submission
Unpublished
2 (bases 1 to 170627)
Worley,K.C.
Direct Submission
Submitted (29-JUN-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 170627)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (09-NOV-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Nov 9, 2002 this sequence version replaced gi:23096537.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rat/) Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GDDP
Center clone name: CH230-9H22
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 152619 bases at least Q40
Consensus quality: 154392 bases at least Q30
Consensus quality: 165289 bases at least Q20
Estimated insert size: 169231; sum-of-contigs estimation
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 4 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 109158: contig of 109158 bp in length
* 109159 109258: gap of unknown length
* 109259 110280: contig of 1022 bp in length
* 110281 110380: gap of unknown length
* 110381 111540: contig of 1160 bp in length
* 111541 111640: gap of unknown length

```

```

FEATURES             * 111641 170627: contig of 58987 bp in length.
Source               1..170627
                    /organism="Rattus norvegicus"
                    /mol_type="genomic DNA"
                    /db_xref="taxon:10116"
                    /clone="CH230-9H22"
misc_feature         1..1284
                    /note="wgs end extension
                    clone_end:Sp6"
misc_feature         3636..4924
                    /note="clone boundary
                    clone_end:Sp6
                    site:EcoRI
BASE COUNT          51711 a 30081 c 30017 g 54086 t 4732 others
ORIGIN
Query Match          9.8%; Score 92.8; DB 2; Length 170627;
Best Local Similarity 48.2%; Pred. NO. 6.1e-05;
Matches 291; Conservative 0; Mismatches 312; Indels 1; Gaps 1;
QY 82 TACCGAATTTTATACCTATTTTGTTCAGATCACTTTATACATAGTGTGGATAGAGTTT 141
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
29284 TAAAAAATATTTTTTCTAAATTTATGTATGTATGTAATAAATAATTTTTTATAGTTAAA 29225
QY 142 TCGATTTAATCAATTAATGTGAACCTTCTGTCACACAGATGTGCATCAGAGGAGTGGT 201
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
29224 TATATATAAAAAATATAAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 29165
QY 202 TTAATAATGAGTAAATTAATGATTAATGATTTAGTCAACGCAACATTTCAAGTTAAG 261
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
29164 ATAAAAAATAATATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 29105
QY 262 AAGTTTTTTCAGAGATACAAAAAGAGTTTCAATTTGAATGATGAAGAAATTTATATTTTA 321
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
29104 AAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 29045
QY 322 AATCATATTTTAAGAGTGAAGTCTTAACGAATCTCATCTAAGAGATGCTAGTGCTCA 381
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
29044 AAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 28985
QY 382 -GAGTTCACCTACTATTTTAACTAAAGCTTTTCAAAAGCTAAAGATTTAAATTTGTT 440
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
28984 TAAATATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 28925
QY 441 ATCAAGAAAGAGAGTTTACAGACGAAAGAACAGTATTTGTTTATGTTTACAGATACACA 500
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
28924 AAATATTTTAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAAT 28865
QY 501 AAAAGCAATATTCAAAACCTGATTTCAGATTAGAGAGATACATTAATAAATAAATCAA 560
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
28864 AATATAAATAAATTTTAAATAATATTTTAAATAAATAAATAAATAAATAAATAAATAA 28805
QY 561 GGTAAATTCGTTTAATCAACTTGAACGATACAAATTTTAAATACGAAGTATTTTATTC 620
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
28804 AAAAAAATATTATTAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 28745
QY 621 AGCATTCGGACATAAATTAATCTAAATTTTAAATTTTGAAGATGCTTTTAAATAGTTA 680
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
28744 AAATATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 28685
QY 681 AAGA 684
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
28684 AAAAA 28681

RESULT 10
PFMALIP3          67970 bp  DNA  linear  INV 29-JAN-2003
LOCUS             Plasmodium falciparum DNA from WAlIP3.
DEFINITION        AL031746 AL844501
ACCESSION          AL031746.9 GI:6594243
VERSION

```

```

KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
AUTHORS
TITLE
JOURNAL
COMMENT
FEATURES
source
gene
CDS
gene
CDS

```

```

Plasmodium falciparum 3D7
Plasmodium falciparum 3D7
Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
1
Hall,N., Pain,A., Berriman,M., Churcher,C., Harris,B., Harris,D.,
Mungall,K., Bowman,S., Atkin,R., Baker,S., Barron,A., Brooks,K.,
Buckee,C.O., Burrows,C., Cherevach,I., Chillingworth,C., Corton,C.,
Chillingworth,T., Christodoulou,Z., Clark,L., Clark,R., Corton,C.,
Cronin,A., Davies,R., Davis,P., Dear,P., Dearden,F., Doggett,J.,
Feltwell,T., Goble,A., Goodhead,I., Gilliam,R., Hamlin,N.,
Hance,Z., Harper,D., Hauser,H., Hornaby,T., Holroyd,S.,
Horrocks,P., Humphray,S., Jagels,K., James,K.D., Johnson,D.,
Kerhornou,A., Knights,A., Konfortov,B., Kyes,S., Larke,N.,
Lawson,D., Lennard,N., Line,A., Maddison,M., McLean,J., Mooney,P.,
Moule,S., Murphy,L., Oliver,K., Ormond,D., Price,C., Quail,M.A.,
Rabbinowitsch,E., Rajandream,M.A., Rutter,S., Rutherford,K.M.,
Sanders,M., Simmonds,M., Seeger,K., Sharp,S., Smith,R., Squares,R.,
Squares,S., Stevens,K., Taylor,K., Tivey,A., Unwin,L.,
Whitehead,S., Woodward,J., Sulston,J.E., Craig,A., Newbold,C. and
Bartell,B.G.
Sequence of Plasmodium falciparum chromosomes 1, 3-9 and 13
Nature 419 (6306), 527-531 (2002)
2225708
12368867
2
Murphy,L., Pain,A., Berriman,M., Bowman,S., Churcher,C., Harris,B.,
Harris,D., Lawson,D., Quail,M., Rajandream,M., Hall,N. and
Bartell,B.
Direct Submission
Submitted (24-SEP-1998) P.falciparum Genome Sequencing Consortium,
The Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge
CB10 1SA, UK
On Dec 16, 1999 this sequence version replaced gi:5763807.
For more information about this sequence or the Malaria Project,
see http://www.sanger.ac.uk/Projects/P_falciparum.
Location/Qualifiers
1..67970
/organism="Plasmodium falciparum 3D7"
/mol_type="genomic DNA"
/isolate="3D7"
/db_xref="taxon:36329"
/chromosome="13"
complement(join(201..257,455..586,748..804))
/gene="PFA0575c"
/note="synonym: MALIP3.01b"
complement(join(201..257,455..586,748..804))
/gene="PFA0575c"
/note="Similar to Schizosaccharomyces pombe hypothetical
protein c119.18 spbc119.18 SWALL:Q96VGI (EMBL:AL022117)
(69 aa) fasta scores: E(): 1.3e-06, 39.68% id in 63 aa,
and to Homo sapiens protein I5el.1 I5el.1 SWALL:I5el1_HUMAN
(SWALL:O43715) (76 aa) fasta scores: E(): 0.00046, 38% id
in 50 aa, and to Mus musculus protein I5el.1 I5el.1
SWALL:Q9D822 (EMBL:AK007514) (76 aa) fasta scores: E():
0.00054, 38% id in 50 aa"
/codon_start=1
/product="hypothetical protein, conserved"
/protein_id="CAD49075.1"
/db_xref="GI:23477009"
/translation="MENRSKTSIQNTIKAHVEANDECKEKEKYLKCFNNWYKNFLK
GDLTQACDDYEDYQICVLNLDNKKGLGHSNVEKEK"
complement(join(1392..1461,1752..2598,2748..2848,
2990..3276))
/gene="PFA0580c"
/note="synonym: MALIP3.01"
complement(join(1392..1461,1752..2598,2748..2848,
2990..3276))
/gene="PFA0580c"
/note="Originally annotated as Plasmodium falciparum;
conserved hypothetical protein but pfam match and other
evidences suggest that it is a putative deoxyribonuclease;
earlier start site possible to make a 412 aa protein

```



```

p1am match to entry PF01026 TatD DNase, TatD related
DNase, score 117.30, E-value 2.9e-31
Similar to Plasmidium falciparum conserved hypothetical
protein, upf0006 family malp3.01 SWALL:Q9U0N6
(EMBL:AL031746) (412 aa) fasta scores: E(): 3.7e-150, 100%
id in 412 aa, and to Saccharomyces cerevisiae putative
deoxyribonuclease yb1055c yb1055c or yb10512 or yb10511
SWALL:YBFS YEAST (SWALL:P34220) (418 aa) fasta scores:
E(): 2.5e-19, 30.47% id in 397 aa
/codon_start=1
/product="tatD-like deoxyribonuclease, putative"
/protein_id="CAD49076.1"
/db_xref="GI:23477010"
/translacion="MKLVFHYKIVNVLVYISIFLKSNSLKIYNDLRYISTVKNKYV
LQIKKRNKLNKHNIRKMDNEDSIFDIGNLTDMPDGVNYSKKHENDLQVNLKAV
NNAKKIIITCTCLAEIDSLKICETVDEPKFLYLGAGVPTNVCYEFIDKNKHEKE
IIAKKEVEERFKYFNKQSVENSKWNGKKICDGEKDMGNLEILLKNDLDTIPGFXY
NEKSKYLENLKNIILKYNRIVCIGIEIGDPLFLYSCVKIOLKVFILKLVQMPN
LPMFLHNCSETFFKVDIYFKFNKGWVHSFTKEDIHVIIVQNYKNLTVGNG
CSLKSLEINAVAKIPLNLLLETDAPWCGVKKTHASYEYIKOTYKRAYTNLKKIKN
IKDDNTIFKREPNYNIADIATYKVEEAAPFDLFCKK"
5005..5496
/gene="MAL1P3.02"
/note="synonym: PPA0585w"
CDS
5005..5496
/gene="MAL1P3.02"
/note="contains possible signal sequence. Signal peptide
predicted SignalP 2.0 HMM (Signal peptide probability
0.640, signal anchor probability 0.357) with cleavage site
probability 0.504 between residues 27 and 28. ScanRegExp
hit to PS00867, Carbamoyl-phosphate synthase subdomain
signature 2"
/codon_start=1
/product="hypothetical protein"
/protein_id="CAB63557.1"
/db_xref="GI:6594245"
/db_xref="SPTREMBL:Q9U0N5"
/translacion="MKLNNRFEVLNCPILIIFFLNSVVLGNRRNNINPHETENAAK
AMRLSGEISIKLDGDELKILNDEKHDKSDKWDKSYFISNLSEKXSQTDLPR
KQKEINEANTKIIEDRQEFYILNDEIENIATRFVLENNFDELYIQFSKLIDIIOS
LNN"
8179..10303
/note="Putative centromere"
14884..20352
/gene="MAL1P3.03"
/note="synonym: PPA0590w"
14884..20352
/gene="MAL1P3.03"
/note="Putative ABC transporter
Pfam match to entry PF00005 ABC tran, ABC transporter,
score 45.00, E-value 1.7e-09; HMMSmart hit to SM00382,
ATPases associated with a variety of cellular activities;
ScanRegExp hit to PS00211, ABC transporters family
signature
Similar to Schizosaccharomyces pombe ATP-binding cassette
transporter abc1 abc1 or spac9e9.12C SWALL:ABCL SCHPO
(SWALL:Q92337) (1427 aa) fasta scores: E(): 4.2e-08,
21.46% id in 1146 aa"
/codon_start=1
/product="ABC transporter, putative"
/protein_id="CAB63558.1"
/db_xref="GI:6594246"
/db_xref="GOA:Q9U0N4"
/db_xref="SPTREMBL:Q9U0N4"
/translacion="WTVYKENVGISNKGKKKSCQNTISFLNLSFDWIRPLINDLIK
GDIQLPNCIRNFDVPYVASKLEENLRLEIVEDSEFYSEKNSNEHLVHHCNSNDASE
KRVYVYHNILWSLTKFKRIILIIIFILETLIVLIGKFDIYVWRLGQKIPV
YISFLKDFVSGLVVYVMMFFLFFLALHPYFHLFTINLKVLSVLYFLYKLNCSNN
NHLQNPDAFYNTVYFSSQTEIDEISRDFLSIGKNSASSSSGKNNKNIIDKNGSVN
DYINIFKSTKMGKSDSLENRSLEFNVINIMFSDVPSVTFVFTSCINLNFVFKIF
MGFVYFHIGKSGVIAIWLIALYLSAMILFEFLPSLFXKGLYLRDKRINDNMHVL
KEFKLKNWNESEFAKINIIPKEMKCYKIRLYLSNIGVFISSISDIVEVIFFI
YLKRLNKEEKFKFTSIMPLYVYKILISVANFENFNNVNMVEGIWIKRLNNVINDH

```

```

LYNDIKNYFMVRYRYNEDYNIWVDKTFLOENITSHDDGTSHNLKHLKNVINKLTLN
MEKYFFVHKMNYHKNINKQLISGLLNVDNTNKKICFOEHKSNSTYNNSSHIHE
KKEVENTHNNSNSWMSNEFEKKEKQNEVILKLENCSEFLSGYDKNCDNDHILKINFN
LARNSLIILGNVSGKSAFPHSILGDFTNMTGNLYIENFKKMPILVVPQSWLFG
NRSMLFGNEVNELLYKYTILQSELLNDLSTIEHGDYKYNDDNLKSGOKVRICLA
RALLYEYIHMKLCIDYEKKLLQPNELDKLINNKNISSNNKSKLVNIPNEN
YFQKCLMDNNFYLLDDIFTDPSISKILFSNFKEDNISKFNKSKLNFADKNCFLSNKS
TLNGLIEDLDNQYEVNIFPIQDTILKIRGNISEYMEKNNLNIKESHWSYSLNT
IDYTRIKLDFEVLNKHVKSNIYKAYFVKGNTEVSFEIDSKIKYIKKMKKNY
KKEHNNKNNNNNNKDDHININMNDNRNNDINLGNSTDSPTVSSIGNE
VLTDTVSNNSDEIKVPLKDTHEEFNKSSMPFVKSSNNINPNFNKYEDNSSS
FKGSIILEYLAFFQOQVPLVTSVIEMLISITDEIKFVPLTWMSIISKXKHSDD
TILQKQXLEYFVILPISLVTSGICFSMIYGNITSAIKVHNNILYSLNAPLYIF
YNNNGNINRIIDISAPDGLKRIYKAFIFPRCISLSSLIILYIMRDCIFPFPV
IILYVFFKRSRQCAQRLYKTSCHTCLONYSALGSKNIILYKNTTHLDVYE
HYNNFRISYFFKLNINWASLYIKIFILLTYTIMPHLYASGIIKLYCKNYVRI
LSTGVCISFSAVLGVIKFLCYDTHIEKMCVQRLBEFAKISKENASNNKENEL
NVIITQYKKNENISDKISAIVEYKNSLSIIINSODDESKYKIGIFENYVYSYK
KIPLVNGYKVIDEBSLKNINMYALKNQKIGIKVGSGAGKSTLLSLGLINISQ
KITVEGRDIRNRGDESLIGLAQSSFFVFNWIRVIFDPYNNFTDDEIVHALKLN
GINLKNLDLYKMHQDKMSYKIKIQSKINQSNNDNTILTLNDCIRYLSVRLYLN
RHKYKILIDEIPIFNLNNSHDELNSFLIGRAKSFNYIIRNHFNNNTVLIISHANT
LSCDDYIVVLRKGEITYRCSEYEDVKTQSELSHLEMD"
23896..26044
/product="18S ribosomal RNA A-type"
/note="18S rRNA expressed in asexual stage parasites"
26045..26438
/product="ITS1 A-type"
/note="ribosomal operon internal transcribed spacer 1
expressed in asexual stage parasites"
26439..26559
/gene="mal1.5.8rRNA"
26439..26559
/gene="mal1.5.8rRNA"
26439..26559
/gene="mal1.5.8rRNA"
/product="5.8S ribosomal RNA"
/note="5.8S ribosomal expressed in asexual stage
parasites
hit to vivax A-type 5.8S ribosomal percent id: 88.00 none"
26560..27435
/product="ITS2 A-type"
/note="ribosomal operon internal transcribed spacer 2
expressed in asexual stage parasites"
27436..31539
/product="28S ribosomal RNA (A-type)"
/note="28S rRNA expressed during asexual development"
complement(join(31566..32476,32675..32775))
/gene="MAL1P3.04"
/note="synonym: PPA0610C"
complement(join(31966..32476,32675..32775))
/gene="MAL1P3.04"
Query Match 9.7%; Score 91.4; DB 3; Length 67970;
Best Local Similarity 47.6%; Pred. No. 0.00013;
Matches 372; Conservative 0; Mismatches 396; Indels 13; Gaps 3;
Cy 50 AAAAAAAAAATCAATAGTGGAGTCATTACCGAATTTTATCTATTATTGTTAG 109
Db 8448 AATATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAA 8507
Cy 110 AATCACTTTTACATAGTGGAGTGGATTTTCGATTTTAAATACATTAAATGACCTT 169
Db 8508 TAAATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 8567
Cy 170 GCTACACAAGATGTGCATGAGGAGTGGTGTAAATAATAGTAAATAATTAATGATTA 229
Db 8568 AATATATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 8627
Cy 230 ATGATTTTAGTCAACCGACATTTCAAGTTAAGAGTTTTTCAGAGATACAAAAAGAGT 289
Db 8628 ATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 8687
Cy 290 TCAATTTGAACATGAAGAAATTTATATTTTAAATCATATTTTAAAGAGTGAGCTTAACG 349
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```



```

CDS
4115..4798
/gene="ATPase6"
/feature="ATP6"
/codon_start=1
/transl_table=5
/product="ATP F0 synthase subunit 6"
/protein_id="AAO18415.1"
/db_xref="GI:27544845"
/translation="MNIMMNLFTDFPSYFYIKFQFNWIFNMMSIIIPLNWLP
SWNLFNKFIFLNFPSFKKKNNINMLIFLSLFFIFNMFNFFSLFYIFSSSTH
LLNLSLWMLSLPLIYINPFLNFKHLVPLNSPKPLMNFVITELSLIRPWT
LSLSLANLSCHLILILLSNFMNWISIIPSMWIQNMLLILSISMSIQAVVFSIL
LSLVFNEN"
4814..5590
/gene="COI11"
4814..5590
/gene="COI11"
/feature="cox3"
/codon_start=1
/transl_table=5
/product="cytochrome c oxidase subunit III"
/protein_id="AAO18416.1"
/db_xref="GI:27544846"
/translation="WKNQFPFLVKNSPRPIINGFNLMNPLISIMMYMKKYLILM
MMSNLIILNSLWFDIIEKNIIPSKSYIKLMLKFSMIMFIISLFFFIFFWAFPS
SISPSIENMWPQMKIENYIPIPLNTLTATSGFFITLSHLVFNLPFSKSIK
LLPFLMLGYSQMIEYKNSYFCINDSIYFSIFLSTSPHGHIIVIGTFMLMTSVI
RYNKHFLSHIHIFELSIWYHFVDVIMFLYLYFYTLI"
5600..5665
/product="tRNA-Gly"
5666..6019
/gene="ND3"
5666..6019
/gene="ND3"
/codon_start=1
/transl_table=5
/product="NADH dehydrogenase subunit 3"
/protein_id="AAO18417.1"
/db_xref="GI:27544847"
/translation="MSMYFMPWAPIMLSYIYNLKYLSMNKINLEKKIPFYGFN
PISKNLPSPSPFLISLLFLIDIEITMLPIMIMLYKLNFLIMLIIFLYMMLI
TLINWMLGYLWMP"
complement(6023..6084)
/product="tRNA-Arg"
6090..6153
/product="tRNA-Asn"
6212..6278
/product="tRNA-Glu"
complement(6273..6337)
/product="tRNA-Phe"
complement(6339..7985)
/gene="ND5"
complement(6339..7985)
/gene="ND5"
/codon_start=1
/transl_table=5
/product="NADH dehydrogenase subunit 5"
/protein_id="AAO18418.1"
/db_xref="GI:27544848"
/translation="MKVLFVSIILFILSFFMLLPGLYFNLVFNELIPFNWIIISNSM
KNMFLINLYKILIVLYVMTSSMYVSYSNLNNFLIKRYFLYLMMLFLLSMVL
ILGNMLITMLGWDGLGTSYCLIIYVKNLSLNSGFTIILNBPGLDGLMIISFLS
AGIYLIYENLFSUNYKXNILLISSITMLISGFNANLEMDFKKIIAPSLUSQGFNM
SLICGLVDLPLHLFIHALPKSNMFCVGFHYMNGSQNFRYKGFYIYPMKSMI
INFSMLMCGFPFLVGVFSKOLMIEIYNNKMSIFCLMILIFSTVLITISYFRIMKQL
LDNNLMCNLIYKVEDKMSYCNMFMILMLFSKPIFNKFLIFNMMLIKIYKFFFK
MFLGYINSLIFDMNFKNLFKNFIYFNWYKFIYNFMNLLNSVEIIFKSFYE
IISNLPFPFSKKIKENKISLYMILMYVLMITFLN"
complement(7986..8043)
/product="tRNA-His"
complement(8044..9366)
/gene="ND4"

CDS
complement(8044..9366)
/gene="ND4"
/codon_start=1
/transl_table=5
/product="NADH dehydrogenase subunit 4"
/protein_id="AAO18419.1"
/db_xref="GI:27544849"
/translation="MYWFMPEFICMFYLFNMYEDLFIILINLXVYLIMIFPFENFM
LWSVMVFNFGNYISNGLIIMLIFIILINLXVYLKYNCLMMNLMMLMYFMYLVP
NENMFLFYFEMFESSLLIFIMYMKWYGBFNSFYLMFTLVFSDLLYLMNML
NLFENSLFYFEMNINLNPNFIMSLVFKIPMYLHGWLLKAHVEAFYFSSMI
LASVMLKLSGYILIRIMYMKMFYLYLMNINFMGLFSLINCLCOLDKLIVAI
SIHMGIMLSMLMLKLSLYSVYMMISHGFISSGFLYINLIVYQNSLKLIFINK
GLINFMPSLMLWEMFCNVSFPFLSNLVSILLMSLIYWKYLFILIIYCVMLSF
IYSMYLFSFVEHGKLFMNFKIYCKMNFSLIMHLPINLNFENLM"
complement(9377..9655)
/gene="ND4L"
complement(9377..9655)
/gene="ND4L"
/codon_start=1
/transl_table=5
/product="NADH dehydrogenase subunit 4L"

Query Match 9.6%; Score 90.6; DB 3; Length 14422;
Best Local Similarity 47.2%; Pred. No. 0.00028;
Matches 345; Conservative 0; Mismatches 379; Indels 7; Gaps 2;

QY 57 AAAAAATCAATAAGTTGGAGTCATTACCGAATTTTATACCTATTTTGTAGAAATGAAC 116
Db 1317 ATAAATTAATAAAGTTAAATATAATAATAATAATAATAATAATAATAATAATAATA 1258
QY 117 TTTATAACATAGTTGGATAGAGTTTTCGATTTAATAACATATAATGTAACCTTGCTACAA 176
Db 1257 AATATAAATATTTAAATTTTATTTTAAATTTTGATAAATTTAAATAATAATAATA 1198
QY 177 CAGATGTGCATCAGAGGAGTGTTTAAATAGTAAATAATAATAATAATAATAATAAT 236
Db 1197 AAATTTTAAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 1138
QY 237 AGTCAACGCAACATTTCAAGTTAAGAAGTTTTCAGAGATACACAAAAGAAAGTTCATTT 296
Db 1137 TAAATATAAAATAATAAAGTTTGTTGTTAAATTTTATAATTTTATAATAAAAGTTTCATTT 1078
QY 297 GAACATGAAGAAATTTATATTTTAAATCATATTTTAAAGAGTGAGTCTAACGAAATCTC 356
Db 1077 ----TATTATAAATCTAGTTTATAAAGGTTATTTCGAATAAATAATAATAATAATA 1022
QY 357 ATCTAAAGAGATTGCTAAGTCTCAGAGTTCAAACCTTACTATTTAACTAAAGCTTTACA 416
Db 1021 ATATATAATTTAAGAAATTTTGAATTAATAATAATAATAATAATAATAATAATAATA 962
QY 417 AAAGCTAAAGATTTTAAATTTGTTTCAAGAAAGAAAGTTTCAAGACGAAAGACAGT 476
Db 961 AAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAAT 902
QY 477 TATGTGTTTGTACAGATACACAAAAGCAATATTTCAAAACTGATTTCAGAATTAGA 536
Db 901 TTATTTTATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 842
QY 537 AGAATACATTAATAATAATAATCAAGGTTAATTCGGTTTAAATCAATGTAACGATAACAAT 596
Db 841 ATAAATTTTTTAAAGAAATAATAATAATAATAATAATAATAATAATAATAATAATAAG 782
QY 597 TTATTAATAGCAAGTTTATTTATTCAGCATTTGGACATAATAATAATAATAATAATAATA 656
Db 781 TTATAAATAATTTATATTTTAAATCAAAATTTTG---TAAATGACAAAAAATGTAATTC 725
QY 657 TTGAAGATGCTTTAAATTAAGTTTAAAGACCAGCCACCTTATTTTACAGCTTTATAAGCTT 716
Db 724 GAATGAATTTTATAAAGTTGATAATAAATAAATAAATAAATAAATAAATAAATAAATTC 665
QY 717 GACAAAGGTACACTAGTCTTTTATTTTAAATATTTTCTTAGAAAATCAAGTTTAGATC 776

```

```

Db      664 AATAAATTATT CATGAATGAATGGAATATTCTTTAATTTTAAAAATATATTTAATAAAA 605
QY.     777 AATAAATTATT 787
LOCUS   777 AATAAATTATT 787
Db      604 AATAAATTATT 594

RESULT 12
PFMAL3P5 86826 bp DNA linear INV 29-JAN-2003
LOCUS Plasmodium falciparum MAL3P5, complete sequence.
DEFINITION AL034556 AL008971 AL008972 AL008978 AL010141 AL010153 AL010162
ACCESSION AL010206 AL010210 AL139179 AL844502
VERSION AL034556.4 GI:234777013
KEYWORDS HTG; centromere; CTRP protein; initiation factor E4;
SOURCE serine/threonine protein phosphatase.
ORGANISM Plasmodium falciparum 3D7
Plasmodium falciparum 3D7
Plasmodium falciparum 3D7
Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE 1 (bases 1 to 86826)
AUTHORS Bowman,S., Lawson,D., Basham,D., Brown,D., Chillingworth,T.,
Churche,C.M., Craig,A., Davies,R.M., Devlin,K., Felcwell,T.,
Gentles,S., Gilliam,R., Hamlin,N., Harris,D., Holroyd,S.,
Hornby,T., Horrocks,P., Jagels,K., Jassal,B., Kyes,S., McLean,J.,
Moule,S., Mungall,K., Murphy,L., Oliver,K., Quail,M.A.,
Rajandream,M.A., Rutter,S., Skelton,J., Squares,R., Squares,S.,
Sulston,J.B., Whitehead,S., Woodward,J.R., Newbold,C. and
Barrell,B.G.
The complete nucleotide sequence of chromosome 3 of Plasmodium
falciparum
NATURE 400 (6744), 532-538 (1999)
JOURNAL 99376085
MEDLINE 10448855
PUBMED
REFERENCE 2
AUTHORS Hall,N., Pain,A., Berriman,M., Churcher,C., Harris,B., Harris,D.,
Mungall,K., Bowman,S., Atkin,R., Baker,S., Barron,A., Brooks,K.,
Buckee,C.O., Burrows,C., Cherevach,I., Chillingworth,C.,
Chillingworth,T., Christodoulou,Z., Clark,L., Clark,R., Corton,C.,
Cronin,A., Davies,R., Davis,P., Dear,P., Dearden,F., Doggett,J.,
Feltwell,T., Goble,A., Goodhead,I., Gwilliam,R., Hamlin,N.,
Hance,Z., Harper,D., Hauser,H., Hornsby,T., Holroyd,S.,
Horrocks,P., Humphray,S., Jagels,K., James,K.D., Johnson,D.,
Kerhornou,A., Knights,A., Konfortov,B., Kyes,S., Larke,N.,
Lawson,D., Lennard,N., Line,A., Maddison,M., McLean,J., Mooney,P.,
Moule,S., Murphy,L., Oliver,K., Ormond,D., Price,C., Quail,M.A.,
Rabinowitz,E., Rajandream,M.A., Rutter,S., Rutherford,K.M.,
Sanders,M., Simmonds,M., Seeger,K., Sharp,S., Smith,R., Squares,R.,
Squares,S., Stevens,K., Taylor,K., Tivey,A., Unwin,L.,
Whitehead,S., Woodward,J., Sulston,J.E., Craig,A., Newbold,C. and
Barrell,B.G.
Sequence of Plasmodium falciparum chromosomes 1, 3-9 and 13
NATURE 419 (6906), 527-531 (2002)
JOURNAL 22255708
MEDLINE 12368867
PUBMED
REFERENCE 3 (bases 1 to 86826)
AUTHORS Lawson,D., Bowman,S. and Barrell,B.
DIRECT SUBMISSION
TITLE Submitted (17-DEC-1998) P.falciparum Genome Sequencing Consortium,
The Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge
CB10 1SA, UK
COMMENT On Oct 2, 2002 this sequence version replaced gi:7711064.
For more information about this sequence or the Malaria Project,
see http://www.sanger.ac.uk/projects/P\_falciparum.
LOCATION/QUALIFIERS
FEATURES
source
1. .86826
/organism="Plasmodium falciparum 3D7"
/mol_type="genomic DNA"
/isolate="3D7"
/db_xref="taxon:36329"
/chromosome="3"
/clones="MAL3P5"
join(322..603,826..1023,1197..1301,1458..1533,1696..2942)
/gene="MAL3P5.1"

/note="synonym: PFC0575w"
join(322..603,826..1023,1197..1301,1458..1533,1696..2942)
/gene="MAL3P5.1"
/note="Possible signal sequence present, revised: added
new exon 2, revised: incorporated new intron in last exon
in comparison with P. knowlesi and P. yoelii"
/codon_start=1
/product="hypothetical protein"
/protein_id="CAB38969.3"
/db_xref="GI:23477014"
/db_xref="SPTREMBL:O97258"
/translation="MTLKNVYIYSSCFILPDLCSFLLKMKYKNHNNKSVTFFL
RSQYRRKFSRKNVSKKKQKPLFLENLKKGFLGFWNQDYQYIDVIS
NINNLTRIKQVTHKSKSEFTNIKQILHCVFSKIDFKIINLISYIKHFQSNIT
VHSILNQISEKVEKKAENYALHLFLDKDENITLFSMMHMDPFKSKQVIEIRD
IKKRIYECVGLIRKKIERYNLYCEKKIKFHKMDAIKKWHEINMKDDDLFVNYHDE
LRCFTMLNIFERNKNLIRSNVDNNDISIDKDMVNNPIDVNNNISLDEKKEO
FNPDDNLKELKDYEQFLPNDNIKIYIEEDQPLYNINSNNNNNNINNTWKX
HKIKOTYNDDDYDIKEEDLVQKNIDDIYKNTIGNKKSLFEPKNOFISQADIEFO
NLSNVNDQHGVRGNDENTKSTEHKNTKNTKNGYTELIOQNMENFKKQINDN
ISDISNNEINIILKLNQSDENILTSDLIYERLTKLWYQIKLYLKFQYDI
INEQPIIKNEKTVLDLNLNGYKIWSPDVDSNLSPEKTKIDSIPEKDKNNQENQKN
SKQYIMFNVM"
complement(3354..6644)
/gene="MAL3P5.2"
/note="MAL3P5.2"
/note="synonym: PFC0580c"
complement(3354..6644)
/gene="MAL3P5.2"
/note="Signal peptide predicted by SignalP 2.0 HMM (Signal
peptide probability 0.635, signal anchor probability 0.287)
with cleavage site probability 0.594 between residues 21
and 22; CDS conserved in P. knowlesi and P. yoelii"
/codon_start=1
/product="hypothetical protein"
/protein_id="CAB38968.1"
/db_xref="GI:4493932"
/db_xref="SPTREMBL:O97257"
/translation="MLGKKRKNVFLVSPSLFAYFLKRHKNNENVETLINNDIE
KIKIRINKCSYIPLLFLNYDSYIYKRLMYLKFGRKGRKDEYIYINNVAK
KREAKYNSDEQNLFNKFIYIYVLEYSKYGLSPHLSYLKKNISCHVCNIYIP
SLYNYKLDNKNLNEKLYFKQINNEHTQOAPTNHTNNNNNNKKPDLINHS
KNTISSYSTNNMKEKNINIDYKNIHNFYTERKSYSDENCOHITLNMILYNLTQ
DNICICANTNIYINFIYMINILKYICYKMEILLNHYHEDMKKKNKNTS
LFKYTFSPFFKKEENIYDLPQDMNHLKENDKFNYSNENTNNIYKYSIDNY
FYDHNSSNRCSFNLAKQOTDNTQHIIMGKPKYPMKSDHEKKNNTGNINIEK
DOKDILKLYFLSKNKLDDQILNELYVIMYMLLPECSKLISIKNHLBKKEG
FDKNKIYILNSADNNLRNLIKFSNBEERENINSFAPFLLSKNILYFEDEIG
RSRDNTIYNNVYDKETNKTNTNNNNNNNDNCSNNDHSCNNNDHSCNNNDHSCN
NNNISCNNNNNICNNNNNICNNNNNICNNMLDEFQCNQDNFNTYRKKRKRRI
VELAKIYTNNIYDYLKGGKKEHQNDNTINLYIKKPPMIFYLKNIINKKDTSFIEH
NNNIWNGDIKNNNIIFKKKYNLFESSIISYFYIKDIYVINYKLYIYDNLKKFCR
VFLKMEHNRKLYKMKAEAPHYLYNFDQFIINNYVHIHKKNIHKLHILKQCKDXE
IDYKFKDLYCMNNINNIIFSYIHKVDHNECVYRIPKAYNKILLIYNNYLNKENYI
YKMKKLYITLNNNINSLIPYINISYKNIYQNKYKNRKNFHSIFSLKNDIHLLEF
LYTQIQNCCDIFSYIYKYNFNKNPLNYLYYELHYIVYSEKKKKKKFFSFSSSP
YSYDTMNSFTFSYFFFSLSYLLFLFYHPDMYASYIFFKTLTYSGLTYSGLYNNI
MVVCGPKT"
join(7669..7717,7800..7829,7912..7940,8064..8106,
8302..8368,8568..8641,8713..8812,8924..9006,9122..9258,
9369..9505,9613..9838)
/gene="PFC0581w"
join(7669..7717,7800..7829,7912..7940,8064..8106,
8302..8368,8568..8641,8713..8812,8924..9006,9122..9258,
9369..9505,9613..9838)
/gene="PFC0581w"
/note="Revised: new gene prediction, splicing prediction
very tentative"
/codon_start=1
/product="hypothetical protein"
/protein_id="CAB90285.1"
/db_xref="GI:7711066"
/db_xref="SPTREMBL:Q9NFA4"
/translation="MKKKKKHYPSIKFVNVVDKKNVLCSDYIRIINMYIGLHIFRL"

```



```

ACCESSION      AE014832 AE014185
VERSION        AE014832.1 GI:23495046
SOURCE         Plasmodium falci-parum 3D7
ORGANISM       Plasmodium falci-parum 3D7
REFERENCE      Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
AUTHORS        Gardner,M.J., Hall,N., Fung,E., White,O., Berriman,M., Hyman,R.W.,
                Carlton,J.M., Pain,A., Nelson,K.E., Bowman,S., Paulsen,I.T.,
                James,K., Eisen,J.A., Rutherford,K., Salzberg,S.L., Craig,A.,
                Kyes,S., Chan,M.-S., Nene,V., Shallow,S.J., Suh,B., Peterson,J.,
                Angiuoli,S., Pertea,M., Allen,J., Selengut,J., Haft,D.,
                Mather,M.W., Vaidya,A.B., Martin,D.M.A., Fairland,A.H.,
                Fraunholz,M.J., Roos,D.S., Ralph,S.A., McFadden,G.I.,
                Cummings,L.M., Subramanian,G.M., Mungall,C., Venter,J.C.,
                Carucci,D.J., Hoffman,S.L., Newbold,C., Davis,R.W., Fraser,C.M. and
                Barrell,B.
TITLE          Genome sequence of the human malaria parasite Plasmodium falci-parum
JOURNAL        Nature 419 (6906), 498-511 (2002)
PUBMED         12368864
FEATURES       2 (bases 1 to 258658)
                Gardner,M.J.
                Direct Submission
                Submitted (13-SEP-2002) The Institute for Genomic Research, 9712
                Medical Center Dr., Rockville, MD 20850, USA
FEATURES       Location/Qualifiers
                1..258658
                /organism="Plasmodium falci-parum 3D7"
                /mol_type="genomic DNA"
                /isolate="3D7"
                /db_xref="taxon:36329"
                /chromosome="10"
                >2..>2542
                /locus_tag="PF10_0181"
                join(<92..2026,2351..>2542)
                /locus_tag="PF10_0181"
                join(92..2026,2351..2542)
                /locus_tag="PF10_0181"
                /codon_start=1
                /product="hypothetical protein"
                /protein_id="AA035379.1"
                /db_xref="GI:23495047"
                /translation="MVHIQESSYNNQKNDKIIICRKNRCRITVHNLVDEPFLSLET
                EDNNLYRYSQMDKKNIDNCLNYKNSGLTRKRNKNKLIKVKKECVHNNI PFSDDLNL
                KKKKHPDYNPFRRKAYSAANTSFDYNNKKKKKKKNNNNKFIYLTNSFONDVNSF
                SHYKKNVKKKEKWDIYKXKTNNDKYKEMKNVKEVLIKDDMQYENKYYDVLNNK
                KIDSLFLSNDNIYRNHVKIYNNLEKVTDPKLPSPENITLNNKTKYQNTIN
                EYVSHNNYNTDEDENKVIKINMNPVYHSTNPSSVTYNNLFSIKLCKACYNLIK
                ISHVEFSQITKYKFDNFNIIICAFLSPEKVEHFYSVD FYKIENLKNKNTSSKS
                KYILFIFKGSVITWENFNHQQKDNFINKILLFNSFSDLLPVLEQEDTIYFFQG
                KSNELRMWKKKKKSLKSLSYEQDKKENNITFVNGVITLNTKSLKLEOKLTASFA
                QSQITQDHEWMDIANTLFIKTSKEIAKGTCTIISKETSRMMNVSSINNVAQ
                FDLPIETVWPKVEHFAWFIYKFEYPERIKILNKNYINIKDFKVIKTEVYNNKT
                FYTIRVMLLLFIHVLAULLNDFLV"
                complement(261..351)
                /note="AT rich"
                /rpt_type=tandem
                477..525
                /rpt_type=tandem
                /rpt_init=(A)n
                581..691
                /rpt_type=tandem
                /rpt_init=(A)n
                complement(732..790)
                /note="AT rich"
                /rpt_type=tandem
                complement(858..914)
                /note="AT rich"
                /rpt_type=tandem
                complement(1209..1246)
                /note="AT rich"
                /rpt_type=tandem

```

```
repeat_region 1477..1623
/rpt_type=tandem
/rpt_unit=(TAAAAA)n
repeat_region 1484..1524
/rpt_type=tandem
/rpt_unit=(A)n
repeat_region 1640..1663
/rpt_type=tandem
/rpt_unit=(A)n
repeat_region 2034..2075
/rpt_type=tandem
/rpt_unit=(A)n
repeat_region 2079..2233
/rpt_type=tandem
/rpt_unit=(A)n
repeat_region 2092..2197
/rpt_type=tandem
/rpt_unit=(A)n
repeat_region 2093..2200
/rpt_type=tandem
/rpt_unit=(TAAAAA)n
repeat_region complement(2222..2273)
/rpt_type=tandem
/rpt_unit=(TA)n
repeat_region 2222..2269
/rpt_type=tandem
/rpt_unit=(A)n
repeat_region complement(2278..2298)
/rpt_type=tandem
/rpt_unit=(A)n
repeat_region complement(2308..2364)
/rpt_type=tandem
/rpt_unit=(TA)n
repeat_region complement(2575..2644)
/rpt_type=tandem
/rpt_unit=(A)n
repeat_region 2575..2628
/rpt_type=tandem
/rpt_unit=(A)n
gene complement(<2637..>4787)
/rpt_type=tandem
/rpt_unit=(A)n
mRNA complement(<2637..>4787)
/rpt_type=tandem
/rpt_unit=(A)n
CDS complement(2637..4787)
/rpt_type=tandem
/rpt_unit=(A)n
/codon_start=1
/product="hypotheical protein"
/protein_id="AA395380.1"
/db_xref="GI:23495048"
/translation="MRIFGSLKRWTVYSPVEYTVILCNEQSRKRNFFSTMLYNTSEL
QNMKNLENNMGLNI SYGIKSKVYGVNKKMFHYNNLIKTYCNFTEDILILQ
SVLCKERNFEVYSLSNKLVIPKDMNEQVNOVDNIYKYLASQOOLNYSYVEIL
TNFLIKNNHLYGLKISILSLTKLKIYDQQLDTSIIYIWNFEHMKNYVSH
IISAYSKQINFLQLYKKLQYIYENIKYMSUSIYNTLIQIRPIIQTVDTNTSLY
KTQDDYCTIQNFKNKSYSSPEDNLFQKKKKKDYDKHNIYNSNSVCDKYIKVELN
NCNSYNNNTDHSNVKRVHTYQSKETYPHTIPVDYTERKTKCTTNNINNDQQKK
NDENWNISFNSSKHSINDEYCHLNDTLKNNKQNNLNKKEKILQNIIPLES
KYNCLAFSPQLVKLACTYKELNVFNIYIKFLPYLANKIQKKISPEECVLL
EGCTVLPVDINIDELIINLTNLTVITFHYNLCRILQCFQHLTLYNDNLISDCI
IIFKNKAFYRCNKLQDLFLHFYKMSDEWQEMINYLKQIISQNNHILQSENSI
QODTHQDSTNNMNMNIYNDMKLLIYKYNKVKQCFNCKEKNKIYEAHTNVHDDKND
ENDGTPVHMENSNDYRKVSQYIYFENTNEQNI"
complement(2821..2848)
/rpt_type=tandem
/rpt_unit=(A)n
repeat_region complement(3335..3365)
/rpt_type=tandem
/rpt_unit=(A)n
repeat_region complement(3453..3531)
/rpt_type=tandem
/rpt_unit=(A)n
repeat_region complement(3861..3880)
/rpt_type=tandem
/rpt_unit=(A)n
repeat_region complement(4042..4111)
/rpt_type=tandem
/rpt_unit=(A)n
```

---

```
complement(4275..4301)
/rpt_type=tandem
/rpt_unit=(A)n
repeat_region complement(4348..4369)
/rpt_type=tandem
/rpt_unit=(A)n
repeat_region complement(4794..4992)
/rpt_type=tandem
/rpt_unit=(A)n
repeat_region complement(4833..4872)
/rpt_type=tandem
/rpt_unit=(TA)n
repeat_region 4837..4872
/rpt_type=tandem
/rpt_unit=(A)n
repeat_region 5124..5152
/rpt_type=tandem
/rpt_unit=(A)n
repeat_region complement(5154..5178)
/rpt_type=tandem
/rpt_unit=(TA)n
repeat_region 5275..5331
/rpt_type=tandem
/rpt_unit=(A)n
repeat_region 5280..5323
/rpt_type=tandem
/rpt_unit=(A)n
repeat_region complement(5332..5357)
/rpt_type=tandem
/rpt_unit=(A)n
repeat_region complement(5378..5402)
/rpt_type=tandem
/rpt_unit=(A)n
repeat_region complement(5495..5532)
/rpt_type=tandem
/rpt_unit=(A)n
repeat_region complement(5578..5604)
/rpt_type=tandem
/rpt_unit=(A)n
repeat_region complement(5670..5689)
/rpt_type=tandem
/rpt_unit=(A)n
repeat_region complement(5749..5838)
/rpt_type=tandem
/rpt_unit=(TA)n
repeat_region 5753..5838
/rpt_type=tandem
/rpt_unit=(A)n
repeat_region complement(5876..5900)
/rpt_type=tandem
/rpt_unit=(A)n
```

Query Match 9.3%; Score 87.6; DB 3; Length 258658;  
Best Local Similarity 47.5%; Pred. No. 0.0003;  
Matches 368; Conservative 0; Mismatches 389; Indels 17; Gaps 3;

Qy 30 AAGTATTGGCGATTTTTCAGAAATAAATAATCAATAAGTTGGAGTCATTACCGAAT 89  
Db 186395 AAATAATATAATAATTTAATTAATAAATAATATATTAACAACAATAAATAATATAAT 186454  
Qy 90 TTTTACTATTGTTTGTAGAAATGAATGAACTTTATACATAGTTGATAGAGTTTCGATTTA 149  
Db 186455 TAATTAATAATAACAATTAATAATTAATAATAATAATAATATTTGTTATATAATTTTA 186514  
Qy 150 ATACATTAAATGTGAACCTTGTCTACACAAGATGTCTACAGAAGAGTGGTTTAAATAAT 209  
Db 186515 ATATATTAAATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 186574  
Qy 210 GAGTAAATTAATGATATTAAATGATTTAGTCAACGCAACATTTCAAGTTAAGAAGTTTT 269  
Db 186575 ATAAATTAATAAT 186634  
Qy 270 CAGAGATACAAAAAGAAGTTCATTTGAACTATGAAGAAATTTATATATTTTAAATCATAT 329





GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 17, 2003, 09:13:34 ; Search time 316 Seconds  
(without alignments)  
8089,772 Million cell updates/sec

Title: US-10-043-539A-1  
Perfect score: 947  
Sequence: 1 gttttcaaatggtggagg.....gtttcataaactgaatcgat 947

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_19Jun03.\*  
1: /SIDSL/gcgdata/Geneseq/Geneseqn-emb1/NA1980.DAT.\*  
2: /SIDSL/gcgdata/Geneseq/Geneseqn-emb1/NA1981.DAT.\*  
3: /SIDSL/gcgdata/Geneseq/Geneseqn-emb1/NA1982.DAT.\*  
4: /SIDSL/gcgdata/Geneseq/Geneseqn-emb1/NA1983.DAT.\*  
5: /SIDSL/gcgdata/Geneseq/Geneseqn-emb1/NA1984.DAT.\*  
6: /SIDSL/gcgdata/Geneseq/Geneseqn-emb1/NA1985.DAT.\*  
7: /SIDSL/gcgdata/Geneseq/Geneseqn-emb1/NA1986.DAT.\*  
8: /SIDSL/gcgdata/Geneseq/Geneseqn-emb1/NA1987.DAT.\*  
9: /SIDSL/gcgdata/Geneseq/Geneseqn-emb1/NA1988.DAT.\*  
10: /SIDSL/gcgdata/Geneseq/Geneseqn-emb1/NA1989.DAT.\*  
11: /SIDSL/gcgdata/Geneseq/Geneseqn-emb1/NA1990.DAT.\*  
12: /SIDSL/gcgdata/Geneseq/Geneseqn-emb1/NA1991.DAT.\*  
13: /SIDSL/gcgdata/Geneseq/Geneseqn-emb1/NA1992.DAT.\*  
14: /SIDSL/gcgdata/Geneseq/Geneseqn-emb1/NA1993.DAT.\*  
15: /SIDSL/gcgdata/Geneseq/Geneseqn-emb1/NA1994.DAT.\*  
16: /SIDSL/gcgdata/Geneseq/Geneseqn-emb1/NA1995.DAT.\*  
17: /SIDSL/gcgdata/Geneseq/Geneseqn-emb1/NA1996.DAT.\*  
18: /SIDSL/gcgdata/Geneseq/Geneseqn-emb1/NA1997.DAT.\*  
19: /SIDSL/gcgdata/Geneseq/Geneseqn-emb1/NA1998.DAT.\*  
20: /SIDSL/gcgdata/Geneseq/Geneseqn-emb1/NA1999.DAT.\*  
21: /SIDSL/gcgdata/Geneseq/Geneseqn-emb1/NA2000.DAT.\*  
22: /SIDSL/gcgdata/Geneseq/Geneseqn-emb1/NA2001A.DAT.\*  
23: /SIDSL/gcgdata/Geneseq/Geneseqn-emb1/NA2001B.DAT.\*  
24: /SIDSL/gcgdata/Geneseq/Geneseqn-emb1/NA2002.DAT.\*  
25: /SIDSL/gcgdata/Geneseq/Geneseqn-emb1/NA2003.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	329	34.7	400	18 AAV75836	Staphylococcus aur
2	261.4	27.6	387	24 ABN92693	Staphylococcus epi
3	96.2	10.2	8056	25 ABZ10246	Haematopoietic cel
4	90	9.5	8056	25 ABZ10246	Haematopoietic cel
5	87.2	9.2	8056	25 ABZ10100	Haematopoietic cel
6	82	8.7	6292	22 AAS46735	Tumour suppressor
7	80.2	8.5	8056	25 ABZ10100	Haematopoietic cel.
8	79.4	8.4	101	18 AAV75910	Staphylococcus aur

C 9	77.6	8.2	867	24	ABN92692	Staphylococcus epi
C 10	75.6	8.0	529	21	AAA66121	E. coli proliferat
C 11	75.2	7.9	12079	24	ABQ76793	pUC19 promoter-ter
C 12	75.2	7.9	12079	24	ABQ74271	Plant specific exp
C 13	75.2	7.9	12085	24	ABQ76792	pUC19 promoter-ter
C 14	75.2	7.9	12085	24	ABV74270	Plant specific exp
C 15	75.2	7.9	12093	24	ABQ76791	pUC19 promoter-ter
C 16	75.2	7.9	12093	24	ABQ74269	Plant specific exp
C 17	75.2	7.9	12241	24	ABQ73049	Tomato anthocyanin
C 18	75.2	7.9	12241	24	AAD36732	Binary vector DNA
C 19	75.2	7.9	13002	24	ABQ76794	pUC19 promoter-ter
C 20	75.2	7.9	13002	24	ABV74272	Plant specific exp
C 21	75.2	7.9	13737	21	AAA54212	Transformation vec
C 22	75.2	7.9	13905	24	ABQ76795	pUC19 promoter-ter
C 23	75.2	7.9	13905	24	ABQ74273	Plant specific exp
C 24	75.2	7.9	14446	24	ASU17548	Plasmid pTM036. S
C 25	75.2	7.9	15430	24	ABQ76796	pB-DHGLA encoding
C 26	75.2	7.9	15430	24	ABV74274	Plant specific exp
C 27	75.2	7.9	17752	24	ABQ76797	pBAR1 encoding de
C 28	75.2	7.9	17752	24	ABV74275	Plant specific exp
C 29	75.2	7.9	19705	24	ABA92074	Vector plasmid pBT
C 30	75.2	7.9	20119	22	AAH26041	Plasmid pNOV524 en
C 31	75.2	7.9	20174	24	ABA92073	Transformation vec
C 32	75.2	7.9	26019	22	AAH26040	Plasmid pNOV523 en
C 33	75.2	7.9	6664	24	AAS61368	Human gene regulat
C 34	74.6	7.9	4086	22	AAH54401	Human epidermidis gen
C 35	74.4	7.9	778	24	ABQ15588	S. epidermidis gen
C 36	74.4	7.9	778	24	ABQ15589	Oligonucleotide fo
C 37	73.2	7.7	6191	24	ABN80161	Human chemically m
C 38	73.2	7.7	6191	24	ABL70282	Human immune syste
C 39	73.2	7.7	6191	24	ABL33217	Signal transductio
C 40	73.2	7.7	6191	24	ABK31307	Human immune syste
C 41	73.2	7.7	15548	24	ABL34155	Human immune syste
C 42	72.2	7.6	3967	22	AAH54106	S. epidermidis gen
C 43	72	7.6	1501	25	ABZ10188	Haematopoietic cel
C 44	72	7.6	2145	21	AAS56364	Escherichia coli f
C 45	72	7.6	4142	22	AAH54879	S. epidermidis gen

ALIGNMENTS

RESULT 1  
AAV75836  
ID AAV75836 standard; DNA; 400 BP.  
XX  
AC AAV75836;  
XX  
DT 16-MAR-1999 (first entry)  
XX  
DE Staphylococcus aureus contig SEQ ID #1525.

XX Computer readable medium; vaccine; S.aureus infection; immunodetection;  
KW cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;  
KW skin infection; surgical wound infection; scalded skin syndrome;  
KW toxic shock syndrome; ds.

XX Staphylococcus aureus.

XX Key Location/Qualifiers

FT misc\_feature 301...360

FT /tag= a

FT /note= "these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"

PR 05-JAN-1996; 96US-0009861.  
 XX (HUMA-) HUMAN GENOME SCI INC.  
 XX Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA;  
 PI Rosen CA;  
 XX WPI; 1997-374922/35.  
 XX Polynucleotide(s) and proteins derived from *Staphylococcus aureus*  
 PT stored on computer readable medium and used in the production of  
 PT anti-S.aureus vaccines  
 XX  
 PS Claim 1; Page 1984-1985; 3271pp; English.  
 XX  
 CC This sequence represents one of 5191 *Staphylococcus aureus* DNA sequences  
 CC of the invention. The DNA sequences are recorded on a computer readable  
 CC medium, preferably selected from a floppy or hard disk, random access  
 CC memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using  
 CC the S.aureus DNA sequences allows putative functions to be assigned so  
 CC that protein-encoding or regulatory regions of commercial, therapeutic or  
 CC industrial importance can be obtained. Specifically, sequences which are  
 CC likely to encode antigens have been identified and these polypeptides can  
 CC be used in a vaccine composition against *S.aureus* infection. The  
 CC polypeptides can also be used in a kit for the immunodetection of  
 CC *S.aureus* in a sample. *S.aureus* is implicated in numerous human diseases,  
 CC including cellulitis, eyelid infections, food poisoning, osteomyelitis,  
 CC skin and surgical wound infections, scalded skin syndrome, toxic shock  
 CC syndrome, etc. Organisms transformed with the DNA sequences can be used  
 CC for recombinant production of the polypeptides. The new DNA sequences  
 CC (and their fragments) are useful as primers or probes for isolating  
 CC homologues of any of the *S.aureus* DNA sequences contained on the  
 CC computer readable medium.  
 XX  
 SQ Sequence 400 BP; 155 A; 36 C; 44 G; 104 T; 61 other;  
 Query Match 34.7%; Score 329; DB 18; Length 400;  
 Best Local Similarity 84.4%; Pred. No. 2.2e-46;  
 Matches 329; Conservative 0; Mismatches 61; Indels 0; Gaps 0;  
 QY 274 GATCAAAAAGAGGTTCAATTTGAACATGAGAAATTTATTTTAAATCATATTTTA 333  
 Db 1 GATCAAAAAGAGGTTCAATTTGAACATGAGAAATTTATTTTAAATCATATTTTA 60  
 QY 334 AGAGTCAGTCTAACGAATCTCATCTAAGAGATTGCTAAGTCTCAGAGTTCAACCT 393  
 Db 61 AGAAGTGAGTCTAACGNAATCTCATCTAAGAGATTGCTAAGTCTCAGAGTTCAACCT 120  
 QY 394 TACTATTAACTAAAGCTTTACAAAAGCTAAAAGATTAAAAATTGTTATCAAGAAAGA 453  
 Db 121 TACTATTAACTAAAGCTTTACAAAAGCTAAAAGATTAAAAATTGTTATCAAGAAAGA 180  
 QY 454 AGTTTACAGACGAAAGACAGTTATTGTTATGTTTACAGATACACAAAAGCAATATT 513  
 Db 181 AGTTTACAGACGAAAGACAGTTATTGTTATGTTTACAGATACACAAAAGCAATATT 240  
 QY 514 CAAAACTGATTTTACAGATTTAGAGATACATTTAAAAATTAAATCAAGGTTAATTCGGT 573  
 Db 241 CAAAACTGATTTTACAGATTTAGAGATACATTTAAAAATTAAATCAAGGTTAATTCGGT 300  
 QY 574 TAATAACATTGAACGATAACAATTATTAATACGAAGTTATTTATTACGATTTGGACAT 633  
 Db 301 NNN 360  
 QY 634 AAAATTAACTTAAATTTAAATTTGAAGA 663  
 Db 361 AAAATTAACTTAAATTTAAATTTGAAGA 390

RESULT 2  
 ABN92693  
 ID ABN92693 standard; DNA; 387 BP.  
 XX

AC ABN92693;  
 XX 24-JUL-2002 (first entry)  
 XX Staphylococcus epidermidis ORF nucleic acid sequence SEQ ID NO.2156.  
 DE Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;  
 XX Staphylococcus epidermidis; gene therapy; gene; ds.  
 KW antibacterial;  
 XX Staphylococcus epidermidis.  
 OS  
 XX US6380370-B1.  
 PN  
 XX 30-APR-2002.  
 PD  
 XX 13-AUG-1998; 98US-0134001.  
 PF  
 XX 14-AUG-1997; 97US-055779P.  
 PR 08-NOV-1997; 97US-064964P.  
 XX (GENO-) GENOME THERAPEUTICS CORP.  
 PA  
 XX Doucette-Stamm LA, Bush D;  
 PI WPI; 2002-381255/41.  
 XX P-PSDB; ABP40148.  
 DR  
 XX Novel isolated nucleic acid encoding a *Staphylococcus epidermidis*  
 PT polypeptide, useful for diagnosing and treating bacterial infections -  
 PT  
 CC Disclosure; SEQ ID 2156; 267pp; English.  
 XX  
 CC ABN90538 to ABN93374 represent *Staphylococcus epidermidis* open reading  
 CC frame (ORF) nucleic acid sequences which encode the amino acid sequences  
 CC given in ABP35124 to ABP37960. The *S. epidermidis* sequences have  
 CC antibacterial activity and can be used in gene therapy. The sequences  
 CC can also be used in the diagnosis and treatment of bacterial infections,  
 CC particularly *S. epidermidis* infections. The sequences can be used to  
 CC screen for compounds able to interfere with the *S. epidermidis* life  
 CC cycle or inhibit *S. epidermidis* infection.  
 CC N.B. The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from the  
 CC USPTO web site.  
 XX  
 SQ Sequence 387 BP; 175 A; 39 C; 55 G; 118 T; 0 other;  
 Query Match 27.6%; Score 261.4; DB 24; Length 387;  
 Best Local Similarity 81.4%; Pred. No. 4.2e-35;  
 Matches 315; Conservative 0; Mismatches 71; Indels 1; Gaps 1;  
 QY 165 ACCTTGCTACACCAAGATGTCATCAGAGGAGTGGTTAATGATGAGTAAATTAATGA 224  
 Db 1 ACCTTGCTACACCAAGATGTCATCAG-AGAGTGGTTTATTAAGGAAATTAAGA 59  
 QY 225 TATTAATGATTTAGTCAACCGCAATTTCAAGTTAAGAGTTTTTTCAGAGATCAAAAA 284  
 Db 60 CATCAATGATTTGTTTAAATGCGACATTTCAAGTAAAAAATTTTTTAGAGATACTAAAA 119  
 QY 285 GAAGTTCAATTTGAACACTATGAGAAATTTATTTTAAATCATATTTTAAAGAGTGATC 344  
 Db 120 GCAATATAATTTAAATATGAGAAATTTATTTCTTAAATCAATTTTGAAGAGTGAATC 179  
 QY 345 TAACGAATCTCATCTAAAGAGATTGCTAAGTGTCTAGAGTTCAAACTTACTATTAAAC 404  
 Db 180 TAACGAATATCTTCAAAAGAAATTTGTCATGTTTCAAGTTTAAACCGTATTATTAAAC 239  
 QY 405 TAAAGCTTTCAAAAGCTAAAGATTTTAAATTTGTTTCAAGAAAGAGAGTTTACAGA 464  
 Db 240 TAAAGCTTTCAAAAGCTAAAGATTTTAAATTTTACTATCTTAAAGAGAGAGCGTGCATGA 299  
 QY 465 CGAAGAGACAGTATTGTTTATGTTTACAGATACACAAAAAGCAAAATTTTCAAAACCTGAT 524  
 Db 300 TGAAAGACAGTATTGTTTATGTTTATGTTATGATGACACGAGAGAAATTAATAAATTTGAT 359

	Matches	378;	Conservative	0;	Mismatches	423;	Indels	8;	Gaps	2;
QY	525	TTTCAAGATTAGAAGAAATACATTAATAA	551							
Dd	360	TTTGAAGATTAGAAGAACTATATCAATA	386							
RESULT 3										
ID	ABZ10246/c									
XX	ID	ABZ10246	standard; DNA; 8056 BP.							
XX	AC	ABZ10246;								
XX	DT	16-JAN-2003	(first entry)							
XX	DE	Haematopoietic cell proliferation disorder related DNA sequence #386.								
XX	KW	Human; haematopoietic cell proliferation disorder; cytostatic;								
XX	KW	gene therapy; lymphocytic leukaemia; acute myelogenous leukaemia;								
XX	KW	cytosine methylation state; gene; ds.								
XX	OS	Homo sapiens.								
XX	XX	WO200277272-A2.								
XX	PN	03-OCT-2002.								
XX	PF	26-MAR-2002; 2002WO-EP03401.								
XX	PR	26-MAR-2001; 2001US-278333P.								
XX	PA	(SPIG-) EPIGENOMICS AG.								
XX	PI	Berlin K, Braun A, Distler J, Guetig D, Howe A, Mueller J;								
XX	PI	Olek A, Pipebrock C, Adorjan P, Grabs G, Lesche R, Leu E,								
XX	PI	Lewin A, Lipscher E, Maier S, Model F, Mueller V, Otto T;								
XX	PI	Pellet C, Schwabe I, Ziesbarth H;								
XX	PT	WPI; 2003-018942/01.								
XX	PT	Detecting and differentiating between hematopoietic cell proliferative disorders, comprises contacting a target nucleic acid with a reagent that distinguishes between methylated and non-methylated CpG dinucleotides -								
XX	PS	Claim 28; SEQ ID 386; 117pp; English.								
CC	CC	The present invention describes a method for detecting and differentiating between haematopoietic cell proliferative disorders associated with at least 1 gene and/or their regulatory regions in a subject. The method comprises contacting a target nucleic acid in a biological sample obtained from the subject with at least 1 reagent, which distinguishes between methylated and non-methylated CpG dinucleotides within the target nucleic acid. ABZ09861 to ABZ1118 represent specifically claimed nucleotide sequences from the present invention. Oligonucleotides from the present invention can be used for differentiating between healthy haematopoietic cells and proliferative disorder haematopoietic cells; for differentiating between acute lymphocytic leukaemia and acute myelogenous leukaemia; as probes for determining the cytosine methylation state and/or single nucleotide polymorphisms (SNPs) of haematopoietic cell proliferation disorder related sequences and their complements; and as primers for the amplification of haematopoietic cell proliferation disorder related DNA sequences. The nucleotide sequences from the present invention can also be used for detecting a predisposition to, differentiation between subclasses, diagnosis, prognosis, treatment and/or monitoring of haematopoietic cell proliferative disorders. The present method enables a highly specific classification of haematopoietic cell proliferative disorders allowing for improved and informed treatment of patients.								
XX	SQ	Sequence 8056 BP; 3711 A; 0 C; 371 G; 3974 T; 0 other;								
Query Match										
Beat Local Similarity 10.2%; Score 96.2; DB 25; Length 8056;										
Pred. No. 1.4e-07;										





CC diseases, by analysing cytosine methylations. The parameters may be  
 CC compared to another set of genetic and/or epigenetic parameters, the  
 CC differences serving as basis for diagnosis and/or prognosis events which  
 CC are disadvantageous to patients. The present sequence is one of the  
 CC 533 genomic sequences derived from tumour suppressor genes and  
 CC oncogenes.

CC Note: The sequence data for this patent did not form part  
 CC of the printed specification, but was obtained in electronic  
 CC format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 6292 BP; 1736 A; 99 C; 1304 G; 3153 T; 0 other;

Query Match 8.7%; Score 82; DB 22; Length 6292;  
 Best Local Similarity 45.9%; Pred. No. 3.4e-05;  
 Matches 350; Conservative 0; Mismatches 410; Indels 2; Gaps 2;

QY 50 AAAAAAATATCATATAGTTGGAGTCATTACCGAATTTTATCTATTGTTAG 109  
 Db 3772 ATACAAAAAATATTAATAAATAAATAAATAAATAAATAAATAAATAA 3713  
 QY 110 AATGAACCTTTATACATAGTTGGATAGATTTTCGATTTAATACATTAAATGTGAACCTT 169  
 Db 3712 AAAAAACCTTTACAAAAACGACAAAAAATATTTTAAATAATAAATAAATAA 3653  
 QY 170 GCTACAAACAGATGTCATCAGAGGAGTGTTTAATATGAGTAAATTAATGATTA 229  
 Db 3652 ATACAAAAAATTAATAAATAAATAAATAAATAAATAAATAAATAAATAA 3593  
 QY 230 ATGATTTAGTCAAGCAACATTTCAAGTTAAGAGTTTTCAGAGATACAAAAAAGAGT 289  
 Db 3592 AAAAAAATTAACGAACACAAAAATAAATAAATAAATAAATAAATAAATAA 3533  
 QY 290 TCAATTTGAATCAGTGAAGAAATTTATATTTTAAATCATATTTTAAAGAGTGAGTCTAACG 349  
 Db 3532 ATTTTAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 3473  
 QY 350 AA-ATCTCATCTAAAGAGATGCTTAAGTCTCAGAGTTCAACCTTACTATTAACTAAA 408  
 Db 3472 AACAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 3413  
 QY 409 GCTTTACAAAAGCTAAAAGATTTAAAATTTGTTATCAAGAAAAGAGTTTACAGACGAA 468  
 Db 3412 CGAAAAATAACGATATAATAAATAAATAAATAAATAAATAAATAAATAAATAA 3353  
 QY 469 AGACAGTTATGTTTATGTTTACAGATACACAAAAACCAATATTCAAAACCTGATTTC 528  
 Db 3352 AAAAAAATTTTAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 3293  
 QY 529 GAATTAGAAGATACATTAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 588  
 Db 3292 GATCGAAAAAATAATCTCTCTATTACTAATTTTAAATACTCTTCAATCCGATCAAA 3233  
 QY 589 ATACAAATTTAATAGAGAGTATTATTCAGCATTTGGGACATAAATAAATAAATAA 648  
 Db 3232 AAAATAATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 3173  
 QY 649 TTTAAATATTGAAGATGTTTAAATAAAGTAAAGACAGCCATACCTTATTTTCAAGTTA 708  
 Db 3172 AAAATAAATCTCGTCCGTTTACCCTAATTAATAAATAAATAAATAAATAAATAA 3114  
 QY 709 TTAGCTTGACAAAGGTACACTAGTCTTTTATTTTAAATATTTTCTTGAATAAATAA 768  
 Db 3113 ACCACCTCCACCTCCGAATTCAAACCAATCTTTTACCTCAACCTTCTTAATAATAA 3054  
 QY 769 TTACGATCATAAATAATTTCTGCGATATAGCTTTTGGATGTT 810  
 Db 3053 TTACAAACATACGCCACGACGCCCACTAATTTTACATTTT 3012

RESULT 7  
 ABZ10100/C  
 ID ABZ10100 standard; DNA; 8056 BP.

XX ABZ10100;  
 AC  
 XX  
 XX  
 DT 16-JAN-2003 (first entry)  
 XX  
 DE Haematopoietic cell proliferation disorder related DNA sequence #240.  
 XX  
 KW Human; haematopoietic cell proliferation disorder; cytostatic;  
 KW gene therapy; lymphocytic leukaemia; acute myelogenous leukaemia;  
 KW cytosine methylation state; gene; ds.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200277272-A2.  
 XX  
 PD 03-OCT-2002.  
 XX  
 PF 26-MAR-2002; 2002WO-EP03401.  
 XX  
 PR 26-MAR-2001; 2001US-278333P.  
 XX  
 (EPIG-) EPIGENOMICS AG.  
 PA  
 XX Berlin K, Braun A, Distler J, Guetig D, Howe A, Mueller J;  
 PI Olek A, Piepenbrock C, Adorjan P, Grabs G, Lesche R, Leu E;  
 PI Lewin A, Lipscher E, Maier S, Model F, Mueller V, Otto T;  
 PI Pelet C, Schwöpe I, Ziebarth H;  
 XX WPI; 2003-018942/01.  
 DR  
 XX Detecting and differentiating between hematopoietic cell proliferative  
 PT disorders, comprising contacting a target nucleic acid with a reagent  
 PT that distinguishes between methylated and non-methylated CpG  
 PT dinucleotides -  
 XX  
 XX Claim 28; SEQ ID 240; 117pp; English.

The present invention describes a method for detecting and differentiating between haematopoietic cell proliferative disorders associated with at least 1 gene and/or their regulatory regions in a subject. The method comprises contacting a target nucleic acid in a biological sample obtained from the subject with at least 1 reagent, which distinguishes between methylated and non-methylated CpG dinucleotides within the target nucleic acid. ABZ09861 to ABZ11118 represent specifically claimed nucleotide sequences from the present invention. Oligonucleotides from the present invention can be used for differentiating between healthy haematopoietic cells and proliferative disorder haematopoietic cells; for differentiating between acute lymphocytic leukaemia and acute myelogenous leukaemia; as probes for determining the cytosine methylation state and/or single nucleotide polymorphisms (SNPs) of haematopoietic cell proliferation disorder related sequences and their complements; and as primers for the amplification of haematopoietic cell proliferation disorder related DNA sequences. The nucleotide sequences from the present invention can also be used for detecting a predisposition to, differentiation between subclasses, diagnosis, prognosis, treatment and/or monitoring of haematopoietic cell proliferative disorders. The present method enables a highly specific classification of haematopoietic cell proliferative disorders allowing for improved and informed treatment of patients.

Sequence 8056 BP; 3711 A; 371 C; 371 G; 3603 T; 0 other;

Query Match 8.5%; Score 80.2; DB 25; Length 8056;  
 Best Local Similarity 45.5%; Pred. No. 6.7e-05;  
 Matches 368; Conservative 0; Mismatches 433; Indels 8; Gaps 2;

QY 46 TTGGAATAAAAAATATCATATAGTTGGAGTTCATTACCGAATTTTATCTATTGTT 105  
 Db 1493 TTTAATAATTAATTTTAAATAAATAAATAAATAAATAAATAAATAAATAA 1434  
 QY 106 TTGAATGAACCTTTATACATAGTTGGATAGATTTTTCGATTTAATACATTAATGAA 165  
 Db 1433 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1374

QY 166 CCTTCTACACAGATGTCATCAGAGGAGTGGTTTAATATGATGAGTAAATTAATGAT 225  
 Db 1373 ATTTTAAATAAAATTAATAATTAATTTTAAATTAATTAATTAATTAATTAATTTT 1314  
 QY 226 ATTAATGATTTAGTCAACGCAACATTTCAAGTTAAGAGTTTTCAGAGATACAAAAG 285  
 Db 1313 TATAATTTATTAATTTTAAATAATTTACGTTTATATATTAATTAATTAATTAATTAAT 1254  
 QY 286 AAGTTCAATTTGAACATGAGAAATTTATATTTTAAATCATATTTTAAAGAGTGTCT 345  
 Db 1253 TTTTATTTTTCGAAAAATTTATTTTATTAATAAAAAATTAATTAATTAATTAATTAAT 1194  
 QY 346 AACGAATCTCATCTAAAGAGATGCTAAGTGCTCAGAGTTCAACCTTACTATTTAACT 405  
 Db 1193 ATTTTAAATAATTAATAATTAATCGAAATATTAATTTTAAATTTTCGATATCGAA 1134  
 QY 406 AAGCTTTCAAAAGCTTAAAGATTTAAATGTTTCAAGAAAGAAAGATTTCAAGAC 465  
 Db 1133 AAAAATATTAATAAACGAAATATATAAGCTTTTAAACGTTTAAAAAAGGATTAATAA 1074  
 QY 466 GAAAGACAGTTATGTTTATGTTACAGATACACAAAAGCAATATTCAAAAACCTGATT 525  
 Db 1073 CGAAATTTTATTTTACGTAATTAATTTTCGAAATTAATAATTAATAATAATTTT- 1015  
 QY 526 TCAGAAATAGAGATACATTAATAATTAATCAAGTTAATTCGTTTAAATCAATGA 585  
 Db 1014 -----TAAATAATTTTAAATAATTTAATAATTAATTAATTTTAAATTAATA 961  
 QY 586 ACGTAAACATTTAATACAGAGTATTTATTCAGCATTTGGACATTAATACTTA 645  
 Db 960 AAAATAAAATTTATTTTAAATAATTTTAAATAATTTTAAATAATTTTAAATA 901  
 QY 646 AAATTTAAATTTGAAGATGCTTTAAATTAAGTTTAAAGACCGACCATACCTTTATTCAGC 705  
 Db 900 TAAATTTAAATAATTCGAAATTTTATACGTAAATAATAATAATAATAATTTT 841  
 QY 706 TTATTAAGCTTGACACAGGTACACTGCTTTTATTTTAAATTTTCTTAGAAATCA 765  
 Db 840 TAATAAAATATTTTAAATAATTCGTTTATAATTTTATTTTAAATTAATAATAAT-A 782  
 QY 766 AGTTTACGATCATAAATTTTCTCGCATATAGCTTTTGGATGTTTCCAAAGTATTTCTCT 825  
 Db 781 ATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTT 722  
 QY 826 ATAATTTGTGCGGATAAGCAAAATCT 854  
 Db 721 AAATTAATTTTAAATAATAATAATAATTTT 693

## RESULT 8

AAV75910/c  
 ID AAV75910 standard; DNA; 101 BP.

XX AC AAV75910;

XX DT 16-MAR-1999 (first entry)

XX DE Staphylococcus aureus contig SEQ ID #1599.

XX KW Computer readable medium; vaccine; S.aureus infection; immunodetection;  
 XX KW cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;  
 XX KW skin infection; surgical wound infection; scalded skin syndrome;  
 XX KW toxic shock syndrome; ds.

XX OS Staphylococcus aureus.

XX PN EP786519-A2.

XX PD 30-JUL-1997.

XX PF 07-JAN-1997; 97EP-0100117.

OS -JAN-1996; 96US-0009861.

(HUYA-) HUMAN GENOME SCI INC.

Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA;  
 Rosen CA;

WPI; 1997-374922/35.

Polynucleotide(s) and proteins derived from *Staphylococcus aureus* -  
 scored on computer readable medium and used in the production of  
 anti-S.aureus vaccines

Claim 1; Page 2007; 3271pp; English.

This sequence represents one of 5191 *Staphylococcus aureus* DNA sequences  
 of the invention. The DNA sequences are recorded on a computer readable  
 medium, preferably selected from a floppy or hard disk, random access  
 memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using  
 the S.aureus DNA sequences allows putative functions to be assigned so  
 that protein-encoding or regulatory regions of commercial, therapeutic or  
 industrial importance can be obtained. Specifically, sequences which are  
 likely to encode antigens have been identified and these polypeptides can  
 be used in a vaccine composition against S.aureus infection. The  
 polypeptides can also be used in a kit for the immunodetection of  
 S.aureus in a sample. S.aureus is implicated in numerous human diseases,  
 including cellulitis, eyelid infections, food poisoning, osteomyelitis,  
 skin and surgical wound infections, scalded skin syndrome, toxic shock  
 syndrome, etc. Organisms transformed with the DNA sequences can be used  
 for recombinant production of the polypeptides. The new DNA sequences  
 (and their fragments) are useful as primers or probes for isolating  
 homologues of any of the S.aureus DNA sequences contained on the  
 computer readable medium.

Sequence 101 BP; 33 A; 20 C; 12 G; 35 T; 1 other;

Query Match 8.4%; Score 79.4; DB 18; Length 101;  
 Best Local Similarity 92.2%; Pred. No. 9.7e-05;

Matches 94; Conservative 0; Mismatches 7; Indels 1; Gaps 1;

QY 96 ACTTATTGTTTGAAGTAACTTTTAAATTAAGTTTAAAGACCGACCATACCTTTATTCAGC 155

Db 101 ACTTATTGTTTGAAGTAACTTTTAAATTAAGTTTAAAGACCGACCATACCTTTATTCAGC 155

QY 156 TAAATGTCACCTTGTCTACACAAAGATGTCATCAGACGAG 197

Db 42 TACATGTGAACCTTGTCTACACAAAGATGTCATCAGACGAG 1

## RESULT 9

ABN92692/c

ID ABN92692 standard; DNA; 867 BP.

XX AC ABN92692;

XX DT 24-JUL-2002 (first entry)

XX DE Staphylococcus epidermidis ORF nucleic acid sequence SEQ ID NO:2155.

XX KW Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;  
 XX KW antibacterial; gene therapy; gene; ds.

XX OS Staphylococcus epidermidis.

XX PN US6380370-B1.

XX PD 30-APR-2002.

XX PF 13-AUG-1998; 98US-0134001.

XX PR 14-AUG-1997; 97US-055779P.

XX PR 08-NOV-1997; 97US-064964P.

XX  
DR  
WPI: 2002-675961/73.



XX New expression cassette for plant genes, useful for preparing  
PT transgenic plants that have increased production of polyunsaturated  
PT fatty acids -  
PS  
XX Example 13; Page 112-121; 188pp; German.  
XX This invention describes novel expression cassette (EC) containing at  
CC least one each of plant promoter (P) and structural gene (SG) expressed  
CC in plants, flanked by specific restriction enzyme (RE) recognition sites.  
CC The EC has the structure (L1-P-SG-L2) where L1 = is a polylinker  
CC (ABQ76798), L2 = any of three synthetic polylinker-terminator-polylinker  
CC sequences reproduced (ABQ76799-ABQ76801) or equivalent RE-site-containing  
CC sequences and n = 1-3. The invention discloses a vector containing this  
CC EC, an organism containing the EC or the vector and a transgenic plant  
CC containing a (non-)functional nucleic acid in the vector. Transgenic  
CC plants e.g. linseed can be prepared with improved production of fatty acid  
CC esters with an increased content of polyunsaturated fatty acids (PUFA),  
CC useful in animal and human nutrition, cosmetics and pharmaceuticals, e.g.  
CC PUFA are known to reduce levels of cholesterol in the blood and to  
CC protect against heart disease. The expression cassettes of the invention  
CC provide increased and more efficient production of fine chemicals  
CC (especially PUFA), including seed-specific production. This sequence  
CC represents a nucleic acid sequence used to illustrate the method of the  
CC invention.  
XX  
SQ Sequence 12079 BP; 2835 A; 3149 C; 3385 G; 2710 T; 0 other;  
Query Match 7.9%; Score 75.2; DB 24; Length 12079;  
Best Local Similarity 86.5%; Pred. No. 0.00045;  
Matches 83; Conservative 0; Mismatches 13; Indels 0; Gaps 0;  
QY 1 GTTTCAAAATCGGTGGAGTGCATGAAAAGTTATTGGGCATTTTTCGAAAATAAAAAA 60  
DB 3749 GTTCTCAAAATCGGTGGAGTGCATGACAAAGTCATCGGGCATTATCTGACATAAACA 3690  
QY 61 ATATCAATAAGTTGGAGTCATTACCGAATTTTATA 96  
DB 3689 CTATCAATAAGTTGGAGTCATTACCCCAATTATGATA 3654  
RESULT 12  
ABV74271/c  
ID ABV74271 standard; DNA; 12079 BP.  
XX  
AC ABV74271;  
XX  
XX 28-MAR-2003 (first entry)  
XX  
DE Plant specific expression vector SEQ ID NO 22.  
XX  
KW Desaturase; fatty acid; ester; polyunsaturated fatty acid; cosmetic;  
KW animal nutrition; pharmaceutical; cholesterol; heart disease; transgenic;  
KW plant; ds.  
XX  
OS Synthetic.  
XX  
XX DE10102337-A1.  
XX  
XX 25-JUL-2002.  
XX  
XX 19-JAN-2001; 2001DE-1002337.  
XX  
XX 19-JAN-2001; 2001DE-1002337.  
XX  
XX (BADI ) BASF PLANT SCI GMBH.  
XX  
XX Lerchl J, Renz A, Heinz E, Domerque F, Zaehlinger U;  
XX WPI; 2002-644810/70.  
XX  
XX Preparing ester containing polyunsaturated fatty acids, useful e.g. in  
XX animal or human nutrition, by transforming organism with desaturase  
PT

PT gene from Phaeodactylum tricornutum -  
XX  
XX Example 11; Page 108-117; 182pp; German.  
XX The invention relates to preparing (M1) fatty acid esters (I) with an  
CC increased content of polyunsaturated fatty acids (II) with at least two  
CC double bonds by introducing into a (I)-producing organism a specific  
CC nucleic acid (A). (M1) is useful for preparing ester containing  
CC polyunsaturated fatty acids. Oils, lipids and (II) or their fractions,  
CC produced by (M1) are used in animal and human nutrition, cosmetics and  
CC pharmaceuticals, e.g. (II) are known to reduce levels of cholesterol in  
CC the blood and to protect against heart disease. Cells that express (A)  
CC are also used to identify (ant)agonists of desaturases, e.g. for  
CC modulating the yield and rate of production of particular fine chemicals  
CC in microorganisms (claimed). Also (A) and derived proteins can be used as  
CC markers of specific genomic regions and in evolutionary/protein  
CC structural studies. (M1) is suitable for large scale production (no known  
CC bioengineering method can produce (II) on a useful scale). The present  
CC sequence is that of a pUC19 based expression vector with a plant specific  
CC promoter and terminator surrounding the site for insertion of an  
CC expression cassette.  
XX  
SQ Sequence 12079 BP; 2835 A; 3149 C; 3385 G; 2710 T; 0 other;  
Query Match 7.9%; Score 75.2; DB 24; Length 12079;  
Best Local Similarity 86.5%; Pred. No. 0.00045;  
Matches 83; Conservative 0; Mismatches 13; Indels 0; Gaps 0;  
QY 1 GTTTCAAAATCGGTGGAGTGCATGAAAAGTTATTGGGCATTTTTCGAAAATAAAAAA 60  
DB 3749 GTTCTCAAAATCGGTGGAGTGCATGACAAAGTCATCGGGCATTATCTGACATAAACA 3690  
QY 61 ATATCAATAAGTTGGAGTCATTACCGAATTTTATA 96  
DB 3689 CTATCAATAAGTTGGAGTCATTACCCCAATTATGATA 3654  
RESULT 13  
ABQ76792/c  
ID ABQ76792 standard; DNA; 12085 BP.  
XX  
AC ABQ76792;  
XX  
XX 25-MAR-2003 (first entry)  
XX  
DE pUC19 promoter-terminator expression cassette SEQ ID 19.  
XX  
KW Promoter; expression cassette; structural gene; plant; transgenic;  
KW linseed; fatty acid ester; polyunsaturated fatty acid; PUFA; cosmetic;  
KW animal nutrition; human nutrition; pharmaceutical; cholesterol; blood;  
KW heart disease; seed-specific; ds.  
XX  
OS Synthetic.  
XX  
XX DE10102338-A1.  
XX  
XX 25-JUL-2002.  
XX  
XX 19-JAN-2001; 2001DE-1002338.  
XX  
XX 19-JAN-2001; 2001DE-1002338.  
XX  
XX (BADI ) BASF PLANT SCI GMBH.  
XX  
XX Lerchl J, Duwenig E, Bischoff F, Heinz E, Drexler H, Scheffler J;  
XX WPI; 2002-675961/73.  
XX  
XX New expression cassette for plant genes, useful for preparing  
XX transgenic plants that have increased production of polyunsaturated  
XX fatty acids -  
XX  
XX Example 13; Page 104-112; 189pp; German.

XX This invention describes novel expression cassette (EC) containing at  
 CC least one each of plant promoter (P) and structural gene (SG) expressed  
 CC in plants, flanked by specific restriction enzyme (RE) recognition sites.  
 CC The EC has the structure (L1-P-SG-L2)<sub>n</sub> where L1 = is a polylinker  
 CC (ABQ76798), L2 = any of three synthetic polylinker-terminator-polylinker  
 CC sequences reproduced (ABQ76799-ABQ76801) or equivalent RE-site-containing  
 CC sequences and n = 1-3. The invention discloses a vector containing this  
 CC EC, an organism containing the EC or the vector and a transgenic plant  
 CC containing a (non-)functional nucleic acid in the vector. Transgenic  
 CC plants e.g. linseed can be prepared with improved production of fatty acid  
 CC esters with an increased content of polyunsaturated fatty acids (PUFA),  
 CC useful in animal and human nutrition, cosmetics and pharmaceuticals, e.g.  
 CC PUFA are known to reduce levels of cholesterol in the blood and to  
 CC protect against heart disease. The expression cassettes of the invention  
 CC provide increased and more efficient production of fine chemicals  
 CC (especially PUFA), including seed-specific production. This sequence  
 CC represents a nucleic acid sequence used to illustrate the method of the  
 CC invention.  
 XX SQ Sequence 12085 BP; 2838 A; 3149 C; 3385 G; 2713 T; 0 other;

Query Match 7.9%; Score 75.2; DB 24; Length 12085;  
 Best Local Similarity 86.5%; Pred. No. 0.00045;  
 Matches 83; Conservative 0; Mismatches 13; Indels 0; Gaps 0;  
 QY 1 GTTTCAAAATCGGTGGAGTGCATGAAAAAGTTATTGGGCATTTTGTAAAAATAAAAA 60  
 DB 3749 GTTCTCAAAATCGGTGGAGTGCATGAAAAAGTCAATCGGCATTTATCTGAACATAAAACA 3690  
 QY 61 ATATCAATAAGTTGGAGTCATACCGAATTTTATA 96  
 DB 3689 CTATCAATAAGTTGGAGTCATACCGAATTTATGATA 3654

RESULT 14  
 ID ABV74270/c  
 XX ABV74270 standard; DNA; 12085 BP.  
 AC ABV74270;  
 XX  
 DT 28-MAR-2003 (first entry)  
 XX  
 DE Plant specific expression vector SEQ ID NO 21.  
 XX  
 KW Desaturase, fatty acid; ester; polyunsaturated fatty acid; cosmetic;  
 KW animal nutrition; pharmaceutical; cholesterol; heart disease; transgenic;  
 KW plant; ds.  
 XX  
 OS Synthetic.  
 XX  
 PN DE10102337-A1.  
 XX  
 PD 25-JUL-2002.  
 XX  
 PF 19-JAN-2001; 2001DE-1002337.  
 XX  
 PR 19-JAN-2001; 2001DE-1002337.  
 XX  
 PA (BADI) BASF PLANT SCI GMBH.  
 XX  
 PI Lerchl J, Renz A, Heinz E, Dörmig F, Zaehring U;  
 XX WPI; 2002-644810/70.  
 XX  
 PT Preparing ester containing polyunsaturated fatty acids, useful e.g. in  
 PT animal or human nutrition, by transforming organism with desaturase  
 PT gene from *Phaeodactylum tricornutum* -  
 XX  
 PS Example 11; Page 100-108; 182pp; German.  
 XX  
 CC The invention relates to preparing (M1) fatty acid esters (I) with an  
 CC increased content of polyunsaturated fatty acids (II) with at least two

CC double bonds by introducing into a (I)-producing organism a specific  
 CC nucleic acid (A) (M1) is useful for preparing ester containing  
 CC polyunsaturated fatty acids. Oils, lipids and (II) or their fractions,  
 CC produced by (M1) are used in animal and human nutrition, cosmetics and  
 CC pharmaceuticals, e.g. (II) are known to reduce levels of cholesterol in  
 CC the blood and to protect against heart disease. Cells that express (A)  
 CC are also used to identify (antagonists of desaturases, e.g. for  
 CC modulating the yield and rate of production of particular fine chemicals  
 CC in microorganisms (claimed). Also (A) and derived proteins can be used as  
 CC markers of specific genomic regions and in evolutionary/protein  
 CC structural studies. (M1) is suitable for large scale production (no known  
 CC bioengineering method can produce (II) on a useful scale). The present  
 CC sequence is that of a pUC19 based expression vector with a plant specific  
 CC promoter and terminator surrounding the site for insertion of an  
 CC expression cassette.  
 XX SQ Sequence 12085 BP; 2838 A; 3149 C; 3385 G; 2713 T; 0 other;

Query Match 7.9%; Score 75.2; DB 24; Length 12085;  
 Best Local Similarity 86.5%; Pred. No. 0.00045;  
 Matches 83; Conservative 0; Mismatches 13; Indels 0; Gaps 0;  
 QY 1 GTTTCAAAATCGGTGGAGTGCATGAAAAAGTTATTGGGCATTTTGTAAAAATAAAAA 60  
 DB 3749 GTTCTCAAAATCGGTGGAGTGCATGAAAAAGTCAATCGGCATTTATCTGAACATAAAACA 3690  
 QY 61 ATATCAATAAGTTGGAGTCATACCGAATTTTATA 96  
 DB 3689 CTATCAATAAGTTGGAGTCATACCGAATTTATGATA 3654

RESULT 15  
 ID ABQ76791/c  
 XX ABQ76791 standard; DNA; 12093 BP.  
 AC ABQ76791;  
 XX  
 DT 25-MAR-2003 (first entry)  
 XX  
 DE pUC19 promoter-terminator expression cassette SEQ ID 18.  
 XX  
 KW Promoter; expression cassette; structural gene; plant; transgenic;  
 KW linseed; fatty acid ester; polyunsaturated fatty acid; PUFA; cosmetic;  
 KW animal nutrition; human nutrition; pharmaceutical; cholesterol; blood;  
 KW heart disease; seed-specific; ds.  
 XX  
 OS Synthetic.  
 XX  
 PN DE10102338-A1.  
 XX  
 PD 25-JUL-2002.  
 XX  
 PF 19-JAN-2001; 2001DE-1002338.  
 XX  
 PR 19-JAN-2001; 2001DE-1002338.  
 XX  
 PA (BADI) BASF PLANT SCI GMBH.  
 XX  
 PI Lerchl J, Duwenig E, Bischoff F, Heinz E, Drexler H, Scheffler J;  
 XX WPI; 2002-675961/73.  
 XX  
 PT New expression cassette for plant genes, useful for preparing  
 PT transgenic plants that have increased production of polyunsaturated  
 PT fatty acids -  
 XX  
 PS Example 13; Page 95-103; 188pp; German.  
 XX

XX This invention describes novel expression cassette (EC) containing at  
 CC least one each of plant promoter (P) and structural gene (SG) expressed  
 CC in plants, flanked by specific restriction enzyme (RE) recognition sites.  
 CC The EC has the structure (L1-P-SG-L2)<sub>n</sub> where L1 = is a polylinker  
 CC (ABQ76798), L2 = any of three synthetic polylinker-terminator-polylinker  
 CC sequences reproduced (ABQ76799-ABQ76801) or equivalent RE-site-containing  
 CC sequences and n = 1-3. The invention discloses a vector containing this  
 CC EC, an organism containing the EC or the vector and a transgenic plant  
 CC containing a (non-)functional nucleic acid in the vector. Transgenic  
 CC plants e.g. linseed can be prepared with improved production of fatty acid  
 CC esters with an increased content of polyunsaturated fatty acids (PUFA),  
 CC useful in animal and human nutrition, cosmetics and pharmaceuticals, e.g.  
 CC PUFA are known to reduce levels of cholesterol in the blood and to  
 CC protect against heart disease. The expression cassettes of the invention  
 CC provide increased and more efficient production of fine chemicals  
 CC (especially PUFA), including seed-specific production. This sequence  
 CC represents a nucleic acid sequence used to illustrate the method of the  
 CC invention.

CC sequences reproduced (ABQ76799-ABQ76801) or equivalent RE-site-containing  
CC sequences and n = 1-3. The invention discloses a vector containing this  
CC EC, an organism containing the EC or the vector, and a transgenic plant  
CC containing a (non-)functional nucleic acid in the vector. Transgenic  
CC plants e.g. linseed can be prepared with improved production of fatty acid  
CC esters with an increased content of polyunsaturated fatty acids (PUFA),  
CC useful in animal and human nutrition, cosmetics and pharmaceuticals, e.g.  
CC PUFA are known to reduce levels of cholesterol in the blood and to  
CC protect against heart disease. The expression cassettes of the invention  
CC provide increased and more efficient production of fine chemicals  
CC (especially PUFA), including seed-specific production. This sequence  
CC represents a nucleic acid sequence used to illustrate the method of the  
XX invention.

SQ Sequence 12093 BP; 2837 A; 3154 C; 3390 G; 2712 T; 0 other;

Query Match 7.9%; Score 75.2; DB 24; Length 12093;  
Best Local Similarity 86.5%; Pred. No. 0.00045;  
Matches 83; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 1 GTTTTCAAAATCGGTGGAGTGCATGAAAAAGTTATTGGGCATTTTTGGAAAAATAAAAA 60  
|||  
Db 3749 GTTCTCAAAATCGGTGGAGTGCATGACAAAGTCATCGGCAATTATCTGACATAAACA 3690

QY 61 ATATCAATAAGTTGGAGTCATTACCGAATTTTATA 96  
|||  
Db 3689 CTATCAATAAGTTGGAGTCATTACCAATTATGATA 3654

Search completed: November 17, 2003, 09:42:18  
Job time : 318 secs

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 17, 2003, 09:13:37 ; Search time 76 Seconds  
(without alignments)  
5499.866 Million cell updates/sec

Title: US-10-043-539A-1  
Perfect score: 947  
Sequence: 1 gtttcaaaatcggtggagg.....gtttcataaactgaatcgat 947

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents, NA: \*  
1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq: \*  
2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq: \*  
3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq: \*  
4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq: \*  
5: /cgn2\_6/ptodata/2/ina/PCTUS\_COMB.seq: \*  
6: /cgn2\_6/ptodata/2/ina/backfiles.seq: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	261.4	27.6	387	4	US-09-134-001C-2156
2	77.6	8.2	867	4	US-09-134-001C-2155
3	77.6	8.2	8763	4	US-09-453-702B-57
4	75.2	7.9	12241	4	US-09-948-138-4
5	75.2	7.9	13737	4	US-09-538-414-10
6	68.6	7.2	447	4	US-09-134-001C-2777
7	67.2	7.1	1349	2	US-08-676-782-10
8	66	7.0	393	1	US-08-248-505-1
9	66	7.0	393	2	US-08-676-782-6
10	65.4	6.9	19124	2	US-08-487-826B-13
11	64.2	6.8	8654	1	US-08-920-812-6
12	64.2	6.8	8654	1	US-08-920-827-6
13	64.2	6.8	8654	1	US-08-921-177-6
14	64.2	6.8	8654	1	US-08-362-577C-6
15	64.2	6.8	8654	2	US-08-920-828-6
16	63	6.7	19124	2	US-08-487-826B-13
17	62.8	6.6	342	2	US-08-676-782-11
18	62.6	6.6	10640	4	US-09-417-485D-5
19	62.4	6.6	6243	2	US-09-056-075-1
20	60.4	6.4	636	3	US-08-998-416-1137
21	60.4	6.4	837	3	US-08-998-416-288
22	59.8	6.3	665	2	US-08-883-795A-36
23	59.4	6.3	20674	4	US-09-641-638-651
24	58.8	6.2	4185	4	US-09-417-485D-7
25	58.2	6.1	450	4	US-09-134-001C-2314
26	57.2	6.0	9048	3	US-08-973-273-4
27	56.8	6.0	615	3	US-08-998-416-186

ALIGNMENTS

RESULT 1

US-09-134-001C-2156  
; Sequence 2156, Application US/09134001C  
; Patent No. 6380370  
; GENERAL INFORMATION:  
; APPLICANT: Lynn Doucette-Stamm et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
; FILE REFERENCE: GTC-007  
; CURRENT APPLICATION NUMBER: US/09134,001C  
; CURRENT FILING DATE: 1998-08-13  
; PRIOR APPLICATION NUMBER: US 60/064,964  
; PRIOR FILING DATE: 1997-11-08  
; PRIOR APPLICATION NUMBER: US 60/055,779  
; PRIOR FILING DATE: 1997-08-14  
; NUMBER OF SEQ ID NOS: 5674  
; SEQ ID NO 2156  
; LENGTH: 387  
; TYPE: DNA  
; ORGANISM: Staphylococcus epidermidis  
US-09-134-001C-2156

Query Match	27.6%	Score 261.4;	DB 4;	Length 387;
Best Local Similarity	81.4%	Pred. No. 8.9e-46;		
Matches	315;	Conservative	0;	Mismatches 71; Indels 1; Gaps 1;
Qy	165	ACCTTGCTACACAGATGTCATCAGAGAGTGGTTTAAATATGATGATTAATGA	224	
Db	1	ACCTTGCTACACAGATGTCATCAG-AGGATGGTTTATATGGAATAAAGA	59	
Qy	225	TATTATGATTAGTCAACGCAATTTCAAGTTAAGAAGTTTTCAGAGATCAAAAA	284	
Db	60	CAICATGATTTGGTTATCGGACATTTCAAGTAAAAAATTTTAGAGATCTAAAA	119	
Qy	285	GAATTCATTTGAATGATGAGAAATTTATTTTAAATCATATTTTAAAGATGAGTC	344	
Db	120	GCAATATAATTTAAATTTATGAAGAAATTTATTTCTTATCATCTTTGAAAAAGTGAATC	179	
Qy	345	TAACGAATCTCTCTAAAGAGATTGCTAAGTGTCTAGAGTTTCAACCTTACTATTAAAC	404	
Db	180	TAACGAATATCTTCAAAAGAAATTTGCTACATGTTAGAGTTTAAACCGTATTATTAAAC	239	
Qy	405	TAAAGCTTTCAAAAGCTAAAAGATTTAAATTTGTTATCAAGAAAGAGATTTTACAGA	464	
Db	240	TAAAGCATTTCAAAATTTAAAGATTTAAATTTACTATCTAAAAAGAGAGCGTGAATGA	299	
Qy	465	CGAAGACACAGTATTGTTTATGTTACAGATACACAAAAAGCAATATTTCAAAACATGAT	524	
Db	300	TGAAAGAACACAGTATTGTTTATGTTATCAGATGAACACGAGAAAAATAAAAAATTTGAT	359	



```

; TYPE: DNA
; ORGANISM: vector, multiple sequences
US-09-948-138-4

Query Match
Best Local Similarity 7.9%; Score 75.2; DB 4; Length 12241;
Matches 83; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 1 GTTTTCAAAATCGGTGGAGTGCATGAAAGATTTATTTGGGCAATTTTGGAAAAATAAAAAA 60
Db 10841 GTTCTCAAAATCGGTGGAGTGCATGACAAAGTCATCGGCAATTTATCTGAACATAAAACA 10900

QY 61 ATATCAATAAGTTGGAGTCATTACCGAATTTTATA 96
Db 10901 CTATCAATAAGTTGGAGTCATTACCAATTATGATA 10936

RESULT 5
US-09-538-414-10/C
; Sequence 10, Application US/09538414
; Patent No. 6346655
; GENERAL INFORMATION:
; APPLICANT: Hohn, T.
; APPLICANT: Salmeron, J.
; APPLICANT: Peters, C.
; APPLICANT: Kendra, D.
; APPLICANT: Reinders, J.
; APPLICANT: Kuznia, R.
; APPLICANT: Dill-Mackey, R.
; TITLE OF INVENTION: Transgenic Plant and Methods
; FILE REFERENCE: sequencelist
; CURRENT APPLICATION NUMBER: US/09/538,414
; CURRENT FILING DATE: 2000-03-29
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatenIn Ver. 2.0
; SEQ ID NO 10
; LENGTH: 13737
; TYPE: DNA
; ORGANISM: Plasmid
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:Plasmid
US-09-538-414-10

Query Match
Best Local Similarity 7.9%; Score 75.2; DB 4; Length 13737;
Matches 83; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 1 GTTTTCAAAATCGGTGGAGTGCATGAAAGATTTATTTGGGCAATTTTGGAAAAATAAAAAA 60
Db 5775 GTTCTCAAAATCGGTGGAGTGCATGACAAAGTCATCGGCAATTTATCTGAACATAAAACA 5716

QY 61 ATATCAATAAGTTGGAGTCATTACCGAATTTTATA 96
Db 5715 CTATCAATAAGTTGGAGTCATTACCAATTATGATA 5680

RESULT 6
US-09-134-001C-2777
; Sequence 2777, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 2777

```

```

; LENGTH: 447
; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-2777

Query Match
Best Local Similarity 7.2%; Score 68.6; DB 4; Length 447;
Matches 179; Conservative 0; Mismatches 184; Indels 0; Gaps 0;

QY 214 AAAATTAATGATATTAATGATTTAGTCAACGCAACATTTCAAGTTAAGAAGTTTTCAGA 273
Db 85 AAAATCAATGACTGCTTTGAAATTTAGCTATGCTGCTATGCTGACAGATTTAAAAGGA 144

QY 274 GATACAAAAAAGAGTTCAATTTGAACCTATGAAGAAATTTATATTTTAAATCATATTTTA 333
Db 145 ATAAATTAATAAAGAAATTTCAATCAGCTTTGAAGAATTTGCAGTATTAATTTATAGT 204

QY 334 AGAAGTGAAGTCTAACGAAATCTCATCTAAAGAGATTTGCTAAGTCTCAGAGTTCAAACT 393
Db 205 GAGATAAAGAGAGAGGAATATTTATCTTAAAGATATTTCAATCACTTAAATTTATAACAA 264

QY 394 TACTATTTAACTAAAGCTTTACAAAAGCTAAAGATTTTAAATTTGTTATCAAGAAAGA 453
Db 265 CCTCAAGTTGTTAAAGCTGTTAAATTTATCTCAAGAAATTTACTTTAATAAAGCGT 324

QY 454 AGTTTACAGACGAAAGAACAGTTATTGTTTATGTTTACAGATACACAAAAGCAATATT 513
Db 325 AATGAACACGATGAAAGAACTGTTTAACTCTTAGTAGATTCTAAGCAGCGTAAAAAATC 384

QY 514 CAAAACTGATTTTCAGAAATTAGAAGATACATTTAAAAATTAATCAAGTTAATTCGTT 573
Db 385 GATGATTTACTTAAACGAGTAAACACCGTATCACAGAACAAATATGAAAAATGAAGTA 444

QY 574 TAA 576
Db 445 TAA 447

RESULT 7
US-08-676-782-10
; Sequence 10, Application US/08676782
; Patent No. 5976792
; GENERAL INFORMATION:
; APPLICANT: CHEUNG, Ambrose
; APPLICANT: FISCHETTI, Vincent A.
; TITLE OF INVENTION: REGULATION OF EXOPROTEIN IN
; TITLE OF INVENTION: STAPHYLOCOCCUS AUREUS
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS, L.L.P.
; STREET: P.O. Box 1404
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States
; ZIP: 22313-1404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/676,782
; FILING DATE: 08-JUL-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/248,505
; FILING DATE: 25-MAY-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: McGowan, Malcolm K.
; REGISTRATION NUMBER: 39,300
; REFERENCE/DOCKET NUMBER: 016921-092
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-6620

```

TELEFAX: (703) 836-2021  
INFORMATION FOR SEQ ID NO: 10:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1349 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-676-782-10

Query Match 7.1%; Score 67.2; DB 2; Length 1349;  
Best Local Similarity 49.2%; Pred. No. 1.4e-05;  
Matches 177; Conservative 0; Mismatches 183; Indels 0; Gaps 0;

QY 207 AATGAGTAAATTAATGATATTAATGATTAGTCAACGCAACATTTCAAGTTAAGAAGTT 266  
DB 874 AATTACAAAATCAATGATGCTTTGAGTTGTTATCAATGTCACCTTATGCTGACAAAT 933  
QY 267 TTTCAGAGATCAAAAAGAAGTTCAATTTGAATATGAAGAAATTTATTTTAAATCA 326  
DB 934 AAAAAAGTTTAATTAAGAAAGGAAATTTCAATAGCTTTGAGAAATTCGCTGATTCACATA 993  
QY 327 TATTTTAAGAAGTGTCTTAACGAAATCTCATCTAAGAGATTCGTAAGTCTCAGAGTT 386  
DB 994 CATCAGCGAACAAGAGAAAGATTAATTTTAAGATATTAATTAATCAATTTAACTA 1053  
QY 387 CAAACCTTACTATTTAACTAAGCTTTTACAAAGCTAAAGATTTAAATTTGTTATCAAA 446  
DB 1054 CAAACACCAACAGTTGTTTAAAGCAGTTTAAATTTTATCTCAAGAAAGATTACTTCGATAA 1113  
QY 447 GAAAGAGATTTTACAGAGCAAGAAAGCAGTTATTTGTTATGTTACAGATACACAAAAGC 506  
DB 1114 AAACGTAATGAGCATGATGAAGAACCTGTTAATTTCTTTAATGCAACAAACGTTAA 1173  
QY 507 AAATATTCAAAACCTGATTTTCAAGATTTAGAGATTAACATTTAAATTTAAATTTAA 566  
DB 1174 AAAAATCGAATCATTTATGAGTCGAGTAAATAAATGAACTCAAGCAACCAACGAAATTA 1233

## RESULT 8

US-08-248-505-1  
Sequence 1, Application US/08248505  
Patent No. 5587288  
GENERAL INFORMATION:  
APPLICANT: CHEUNG, Ambrose  
APPLICANT: FISCHETTI, Vincent A.  
TITLE OF INVENTION: REGULATION OF EXOPROTEIN IN  
TITLE OF INVENTION: STAPHYLOCOCCUS AUREUS  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESS: BURNS, DOANE, Swecker & Mathis  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/248,505  
FILING DATE: 24-MAY-1994  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Crane-Feury, Sharon E  
REGISTRATION NUMBER: 36,113  
REFERENCE/DOCKET NUMBER: 016921-018  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-2021  
INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:  
LENGTH: 393 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 19..390  
US-08-248-505-1

Query Match 7.0%; Score 66; DB 1; Length 393;  
Best Local Similarity 48.6%; Pred. No. 2.1e-05;  
Matches 180; Conservative 0; Mismatches 190; Indels 0; Gaps 0;

QY 207 AATGAGTAAATTAATGATATTAATGATTAGTCAACGCAACATTTCAAGTTAAGAAGTT 266  
DB 24 AATTACAAAATCAATGATGCTTTGAGTTGTTATCAATGTCACCTTATGCTGACAAAT 83  
QY 267 TTTCAGAGATCAAAAAGAAGTTCAATTTGAATATGAAGAAATTTATTTTAAATCA 326  
DB 84 AAAAAAGTTTAAATTAAGAAAGGAAATTTCAATAGCTTTGAGAAATTCGCTGATTCACATA 143  
QY 327 TATTTTAAGAAGTGTCTTAACGAAATCTCATCTAAGAGATTCGTAAGTCTCAGAGTT 386  
DB 144 CATCAGCGAACAAGAGAAAGATTAATTTTAAGATATTAATTAATCAATTTAACTA 203  
QY 387 CAAACCTTACTATTTAACTAAGCTTTTACAAAAGCTTAAAGATTTTAAATTTGTTATCAAA 446  
DB 204 CAAACACCAACAGTTGTTTAAAGCAGTTTAAATTTTATCTCAAGAAAGATTACTTCGATAA 263  
QY 447 GAAAGAGATTTTACAGAGCAAGAAAGCAGTTATTTGTTATGTTACAGATACACAAAAGC 506  
DB 264 AAACGTAATGAGCATGATGAAGAACCTGTTAATTTCTTTAATGCAACAAACGTTAA 323  
QY 507 AAATATTCAAAACCTGATTTTCAAGATTTAGAGATTAACATTTAAATTTAAATTTAA 566  
DB 324 AAAAATCGAATCATTTATGAGTCGAGTAAATAAAGCAATCACTGAAGCAACGAAAT 383  
QY 567 TTGCGTTTAA 576  
DB 384 TGAACATATAA 393

## RESULT 9

US-08-676-782-6  
Sequence 6, Application US/08676782  
Patent No. 5976792  
GENERAL INFORMATION:  
APPLICANT: CHEUNG, Ambrose  
APPLICANT: FISCHETTI, Vincent A.  
TITLE OF INVENTION: REGULATION OF EXOPROTEIN IN  
TITLE OF INVENTION: STAPHYLOCOCCUS AUREUS  
NUMBER OF SEQUENCES: 18  
CORRESPONDENCE ADDRESS:  
ADDRESS: BURNS, DOANE, Swecker & Mathis, L.L.P.  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/676,782  
FILING DATE: 08-JUL-1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/248,505  
FILING DATE: 25-MAY-1994





Db 15938 TAATTAATAAATAAATAAATAAATAAATAAATAAATACATGC 15976

## RESULT 11

US-08-920-812-6/c  
; Sequence 6, Application US/08920812  
; Patent No. 5763188  
; GENERAL INFORMATION:  
; APPLICANT: Ohno, Tsuneya  
; APPLICANT: Matsuhisa, Akio  
; APPLICANT: Uehara, Hirotugu  
; APPLICANT: Eda, Soji  
; TITLE OF INVENTION: Probe for Diagnosing Infectious Disease  
; NUMBER OF SEQUENCES: 25  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
; STREET: 6300 Sears Tower, 233 South Wacker Drive  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: United States of America  
; ZIP: 60606-6402  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION NUMBER: US/08/920,812  
; FILING DATE: 29-AUG-1997  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/362,577  
; FILING DATE: 27-MAR-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Rin-Laures, Li-Hsien  
; REGISTRATION NUMBER: 33,547  
; REFERENCE/DOCKET NUMBER: 19036/32420  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 312/474-6300  
; TELEFAX: 312/474-0448  
; TELEX: 25-3856  
; INFORMATION FOR SEQ ID NO: 6:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 8654 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: Genomic DNA  
; ORIGINAL SOURCE:  
; ORGANISM: Staphylococcus epidermidis  
; STRAIN: Clinical Isolate SE-22

## US-08-920-812-6

Query Match 6.8%; Score 64.2; DB 1; Length 8654;  
Best Local Similarity 49.0%; Pred. No. 8e-05;  
Matches 171; Conservative 0; Mismatches 178; Indels 0; Gaps 0;  
QY 214 AAAATTAATGATTAATGATTTAGTCAACGCAACATTTCAAGTTAAGAAGTTTTCAGA 273  
Db 1938 AAAATCAATGACTGCTTTGAAATTAATAGCTATGCTGACAGATTAAAGGA 1879  
QY 274 GATACAAAAAAGAAGTTCAATTTGAAGAAATTTATATTTAAATCATATTTTA 333  
Db 1878 AFAATTAATAAAGAATTTCTCAATCAGCTTGAAGAAATTTGAGTATTAACTATATTAGT 1819  
QY 334 AGAAGTGACTTAAGAAATCTCATCTAAGAGATTGCTAGTCTGACAGTTCAACCT 393  
Db 1818 GAGATAAAGAAGAGGAATATTATCTTAAAGATATATCAATCACTTAAATATATAACAA 1759  
QY 394 TACTATTTAACTAAAGCTTTACAAAAGCTTAAAGAAATTTAAATTTGTTATCAAGAAAGA 453  
Db 1758 CCTCAAGTTGTTAAGCTGTTAAATTTATCTCAAGAAATTTACTTTTAAATAAAGCGT 1699

QY 454 AGTTTCAAGACGAAAGAACAGTATTGTTTGTACAGATACACAAAAAGCAATATT 513  
Db 1698 AATGAACACGATGAAGAACTGTTTAACTCTTAGTAGATTCTAAGCAGCGTAAAAAAATC 1639  
QY 514 CAAAAAAGTATTTCAAGAAATTAAGAGATACATTAATAAATTAATCAAGG 562  
Db 1638 GATGATTTTACTTAACGAGTAACCAACCGCTATCACAAGAGCAATAATG 1590

## RESULT 12

US-08-920-827-6/c  
; Sequence 6, Application US/08920827  
; Patent No. 5770375  
; GENERAL INFORMATION:  
; APPLICANT: Ohno, Tsuneya  
; APPLICANT: Matsuhisa, Akio  
; APPLICANT: Uehara, Hirotugu  
; APPLICANT: Eda, Soji  
; TITLE OF INVENTION: Probe for Diagnosing Infectious Disease  
; NUMBER OF SEQUENCES: 25  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
; STREET: 6300 Sears Tower, 233 South Wacker Drive  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: United States of America  
; ZIP: 60606-6402  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION NUMBER: US/08/920,827  
; FILING DATE: 29-AUG-1997  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/362,577  
; FILING DATE: 27-MAR-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Rin-Laures, Li-Hsien  
; REGISTRATION NUMBER: 33,547  
; REFERENCE/DOCKET NUMBER: 19036/32420  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 312/474-6300  
; TELEFAX: 312/474-0448  
; TELEX: 25-3856  
; INFORMATION FOR SEQ ID NO: 6:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 8654 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: Genomic DNA  
; ORIGINAL SOURCE:  
; ORGANISM: Staphylococcus epidermidis  
; STRAIN: Clinical Isolate SE-22

## US-08-920-827-6

Query Match 6.8%; Score 64.2; DB 1; Length 8654;  
Best Local Similarity 49.0%; Pred. No. 8e-05;  
Matches 171; Conservative 0; Mismatches 178; Indels 0; Gaps 0;  
QY 214 AAAATTAATGATTAATGATTTAGTCAACGCAACATTTCAAGTTAAGAAGTTTTCAGA 273  
Db 1938 AAAATCAATGACTGCTTTGAAATTAATAGCTATGCTGACAGATTAAAGGA 1879  
QY 274 GATACAAAAAAGAAGTTCAATTTGAAGAAATTTATATTTAAATCATATTTTA 333  
Db 1878 ATAATTAATAAAGAATTTCTCAATCAGCTTGAAGAAATTTGAGTATTAACTATATTAGT 1819  
QY 334 AGAAGTGACTTCAACGAAATCTCATCTAAGAGATTGCTAGTCTGACAGTTCAACCT 393

Db 1818 GAGATAAAGAGAGGAAATATTATCTTAAAGATATTATCAATCACTTAAATTTATAAACAA 1759  
QY 394 TACTATTAACTAAGCTTTACAAAGCTAAAGATTTAAATTTGTTATCAAGAAAGA 453  
Db 1758 CCTCAAGTTGTTAAGCTGTTTAAATTTATCTCAAGAAATTTACTTTAATAAAAAACGT 1699  
QY 454 AGTTTACAGACGAAAGACAGTTATTGTTTATGTTACAGATACACAAAAAGCAAAATATT 513  
Db 1698 AATGAACACGATGAAGAACTGTTTAAATCTTAGTAGATCTTAGCAGCGTAAAAAAATC 1639  
QY 514 CAATAACTGATTTCAAGATTAGAGATACATTTAAATTTAAATCAAGG 562  
Db 1638 GATGATTTACTTAAACGAGTAAACACCGTATACAGAGCAAAATAATG 1590  
RESULT 13  
US-08-921-177-6/C  
; Sequence 6, Application US/08921177  
; Patent No. 5798211  
; GENERAL INFORMATION:  
; APPLICANT: Ohno, Tsuneya  
; APPLICANT: Matsuhisa, Akio  
; APPLICANT: Uehara, Hirotugu  
; APPLICANT: Eda, Soji  
; TITLE OF INVENTION: Probe for Diagnosing Infectious Disease  
; NUMBER OF SEQUENCES: 25  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
; STREET: 6300 Sears Tower, 233 South Wacker Drive  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: United States of America  
; ZIP: 60606-6402  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/921.177  
; FILING DATE: 29-AUG-1997  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/362,577  
; FILING DATE: 27-MAR-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Rin-Laures, Li-Hsien  
; REGISTRATION NUMBER: 33,547  
; REFERENCE/DOCKET NUMBER: 19036/32420  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 312/474-6300  
; TELEFAX: 312/474-0448  
; TELEX: 25-3856  
; INFORMATION FOR SEQ ID NO: 6:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 8654 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: Genomic DNA  
; ORIGINAL SOURCE:  
; ORGANISM: Staphylococcus epidermidis  
; STRAIN: Clinical Isolate SE-22  
US-08-921-177-6

Query Match 6.8%; Score 64.2; DB 1; Length 8654;  
Best Local Similarity 49.0%; Pred. No. 8e-05;  
Matches 171; Conservative 0; Mismatches 178; Indels 0; Gaps 0;  
QY 214 AAATTAATGATTAATTAATGATTTAGTATGTTAGTCAACGCAACATTTCAAGTTAAGAGTTTTCAGA 273  
Db 1938 AAAATCAATGATGCTGCTTTGAATTTAGTATGTTAGTCAACGCAACATTTCAAGTTAAGAGTTTTCAGA 1879

QY 274 GATACAAAAAGAGTTCAATTTGAACATATGAAGAAATTTATATTTTAAATCATATTITA 333  
Db 1878 ATAAATTAAGAAAGATTTCTCAATCAGCTTTGAAGATTTGCGAGTATTAATTAATAGT 1819  
QY 334 AGAAGTGAGTCTAACGAATCTCACTAAAGAGATTGCTAAGTGTCTGAGAGTTCAAACT 393  
Db 1818 GAGATAAAGAGAGGAAATATTATCTTAAAGATATTATCAATCACTTAAATTTATAAACAA 1759  
QY 394 TACTATTTAACTAAAGCTTTACAAAAGCTAAAGATTTAAATTTGTTATCAAGAAAGA 453  
Db 1758 CCTCAAGTTGTTAAGCTGTTTAAATTTATCTCAAGAAATTTACTTTAATAAAAAACGT 1699  
QY 454 AGTTTACAGACGAAAGACAGTTATTGTTTATGTTTACAGATACACAAAAAGCAAAATATT 513  
Db 1698 AATGAACACGATGAAGAACTGTTTAAATCTTAGTAGATCTTAAAGCAGCGTAAAAAAATC 1639  
QY 514 CAATAACTGATTTCAAGATTAGAGATACATTTAAATTTAAATCAAGG 562  
Db 1638 GATGATTTACTTAAACGAGTAAACACCGTATACAGAGCAAAATAATG 1590  
RESULT 14  
US-08-362-577C-6/C  
; Sequence 6, Application US/08362577C  
; Patent No. 5807673  
; GENERAL INFORMATION:  
; APPLICANT: Ohno, Tsuneya  
; APPLICANT: Matsuhisa, Akio  
; APPLICANT: Uehara, Hirotugu  
; APPLICANT: Eda, Soji  
; TITLE OF INVENTION: Probe for Diagnosing Infectious Disease  
; NUMBER OF SEQUENCES: 25  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
; STREET: 6300 Sears Tower, 233 South Wacker Drive  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: United States of America  
; ZIP: 60606-6402  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/362,577C  
; FILING DATE: 27-MAR-1995  
; CLASSIFICATION: 536  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Rin-Laures, Li-Hsien  
; REGISTRATION NUMBER: 33,547  
; REFERENCE/DOCKET NUMBER: 19036/32420  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 312/474-6300  
; TELEFAX: 312/474-0448  
; TELEX: 25-3856  
; INFORMATION FOR SEQ ID NO: 6:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 8654 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: Genomic DNA  
; ORIGINAL SOURCE:  
; ORGANISM: Staphylococcus epidermidis  
; STRAIN: Clinical Isolate SE-22  
US-08-362-577C-6

Query Match 6.8%; Score 64.2; DB 1; Length 8654;  
Best Local Similarity 49.0%; Pred. No. 8e-05;  
Matches 171; Conservative 0; Mismatches 178; Indels 0; Gaps 0;  
QY 214 AAATTAATGATTAATTAATGATTTAGTATGTTAGTCAACGCAACATTTCAAGTTAAGAGTTTTCAGA 273

```

Db 1938 AAAATCAATGACTGCTTTGAATTATTAGCTATGCTCACTTATGCTGACAGATTAAAGGA 1879
QY 274 GATACAAAAGAAAGATTCAATTTGAACCTATGAAGAAATTTATATTTAAATCATATTTTA 333
Db 1878 ATAATTTAAAGAAAGATTCTCAATCAGCTTTGAAGAAATTTGCAAGTATTAACCTATATTAGT 1819
QY 334 AGAAGTGAGTCTAACGAAATCTCATCTAAAGAGATTGCTAAGTGCTCAGAGTTCAAACT 393
Db 1818 GAGATTAAGAGAGAGATATATCTTAAAGATATATCAATCACTTTAAATTAACAA 1759
QY 394 TACTATTTAACTAAAGCTTTACAAAGCTAAAGATTTAAATTTGTTATCAAGAAAGA 453
Db 1758 CCTCAAGTTGTTAAAGCTGTTAAAGAAATTTATCTCAAGAAATTTACTTTAATAAAAAACGT 1699
QY 454 AGTTTACAGACGAAAGACAGTATTGTTTATGTTTACAGATACACAAAAGCAATATT 513
Db 1698 AATGAACACGATGAAGAACTGTTTAAATCTTAGTAGATTCTAAGCAGCGTAAAAAAATC 1639
QY 514 CAAAACTGATTTACAGAAATTAGAAGATACATTTAAAAATTTAAATCAAGG 562
Db 1638 GATGATTTACTTTAAACGAGTAAACCAACCGTATCACAGAAGCAATAATG 1590

```

## RESULT 15

```

US-08-920-828-6/c
; Sequence 6, Application US/08920828
; Patent No. 5853998
; GENERAL INFORMATION:
; APPLICANT: Chno, Tsuneya
; APPLICANT: Matsuhisa, Akio
; APPLICANT: Uehara, Hirotsugu
; APPLICANT: Eda, Soji
; TITLE OF INVENTION: Probe for Diagnosing Infectious Disease
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 6300 Sears Tower, 233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/920,828
; FILING DATE: 29-AUG-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/362,577
; FILING DATE: 27-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Rin-Laures, Li-Hsien
; REGISTRATION NUMBER: 33,547
; REFERENCE/DOCKET NUMBER: 19036/32420
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312/474-6300
; TELEFAX: 312/474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8654 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; ORIGINAL SOURCE:
; ORGANISM: Staphylococcus epidermidis
; STRAIN: Clinical Isolate SE-22
US-08-920-828-6

```

```

Query Match 6.8%; Score 64.2; DB 2; Length 8654;
Best Local Similarity 49.0%; Pred. No. 8e-05;
Matches 171; Conservative 0; Mismatches 178; Indels 0; Gaps 0;
QY 214 AAAATTAATGATATTAATGATTTAGTCAACGCAACATTTCAAGTTAAGAAGTTTTTCAGA 273
Db 1938 AAAATCAATGACTGCTTTGAATTATTAGCTATGCTCACTTATGCTGACAGATTAAAGGA 1879
QY 274 GATACAAAAGAAAGTTCAATTTGAACCTATGAAGAAATTTATATTTAAATCATATTTTA 333
Db 1878 ATAATTTAAAGAAAGATTTCTCAATCAGCTTTGAAGAAATTTGCAAGTATTAACCTATATTAGT 1819
QY 334 AGAAGTGAGTCTAACGAAATCTCATCTAAAGAGATTGCTAAGTGCTCAGAGTTCAAACT 393
Db 1818 GAGATTAAGAGAGAGATATATCTTAAAGATATATCAATCACTTTAAATTAACAA 1759
QY 394 TACTATTTAACTAAAGCTTTACAAAGCTAAAGATTTAAATTTGTTATCAAGAAAGA 453
Db 1758 CCTCAAGTTGTTAAAGCTGTTAAAGAAATTTATCTCAAGAAATTTACTTTAATAAAAAACGT 1699
QY 454 AGTTTACAGACGAAAGACAGTATTGTTTATGTTTACAGATACACAAAAGCAATATT 513
Db 1698 AATGAACACGATGAAGAACTGTTTAAATCTTAGTAGATTCTAAGCAGCGTAAAAAAATC 1639
QY 514 CAAAACTGATTTACAGAAATTAGAAGATACATTTAAAAATTTAAATCAAGG 562
Db 1638 GATGATTTACTTTAAACGAGTAAACCAACCGTATCACAGAAGCAATAATG 1590

```

Search completed: November 17, 2003, 11:25:43

Job time : 79 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 17, 2003, 10:44:07 ; Search time 363 Seconds  
(without alignments)  
8526.142 Million cell updates/sec

Title: US-10-043-539A-1

Perfect score: 947

Sequence: 1 GTTTCAAAATCGTGAGGTGCATGAAAGTATTGGCATTTCGAAATATAAAAA 60

Scoring table: IDENTITY NUC

Gapop 10\*0, Gapext 1.0

Searched: 2169951 seqs, 1634102185 residues

Total number of hits satisfying chosen parameters: 4339922

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA:\*

```
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq.*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq.*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq.*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq.*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq.*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq.*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq.*
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq.*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq.*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq.*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq.*
13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq.*
14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq.*
15: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq.*
16: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq.*
17: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	947	100.0	947	14	US-10-043-539-1
2	387.4	40.9	400	8	US-08-781-986A-1525
3	210	22.2	210	14	US-10-043-539-19
C 4	82.4	8.7	3673778	12	US-10-312-841-1
C 5	79.4	8.4	101	8	US-08-781-986A-1599
C 6	77.6	8.2	87563	14	US-10-114-170-57
C 7	76.6	8.1	3252	12	US-10-027-632-113786
8	75.6	8.1	3252	12	US-10-027-632-113787
9	75.6	8.1	3252	12	US-10-027-632-113788
10	75.6	8.1	3252	13	US-10-027-632-113786
11	75.6	8.1	3252	13	US-10-027-632-113787
12	75.6	8.1	3252	13	US-10-027-632-113788
C 13	75.6	8.0	529	9	US-09-912-020-468
C 14	75.2	7.9	11777	12	US-10-122-085A-5
C 15	75.2	7.9	13241	13	US-10-033-190-5
C 16	75.2	7.9	13737	13	US-10-074-279-10

```
C 17 75.2 7.9 14108 12 US-09-845-064-56
C 18 75.2 7.9 14446 10 US-09-810-861B-4
C 19 75.2 7.9 15077 12 US-09-845-064-57
C 20 75.2 7.9 15208 12 US-09-845-064-51
C 21 75.2 7.9 20119 12 US-10-148-907A-35
C 22 75.2 7.9 26019 12 US-10-148-907A-34
C 23 74.6 7.9 3673778 12 US-10-312-841-2
C 24 73.2 7.7 6191 12 US-10-311-455-1190
C 25 73.2 7.7 15548 12 US-10-311-455-2128
C 26 71.4 7.5 6061 12 US-10-311-455-114
C 27 71 7.5 1398 8 US-08-781-986A-253
C 28 71 7.5 7857 12 US-10-311-455-1995
C 29 71 7.5 7657 14 US-10-239-676-185
C 30 70.6 7.5 12237 12 US-10-311-455-2331
C 31 70.4 7.4 6255 12 US-10-311-455-933
C 32 70.4 7.4 17458 14 US-10-055-001A-25
C 33 70.4 7.4 17476 12 US-10-385-546-7
C 34 70.4 7.4 17476 14 US-10-055-001A-24
C 35 70.4 7.4 17881 14 US-10-055-001A-26
C 36 70.4 7.4 17862 14 US-10-055-001A-23
C 37 70.4 7.4 18691 14 US-10-055-001A-13
C 38 70.4 7.4 3673778 12 US-10-312-841-1
C 39 69.8 7.4 5296 12 US-10-311-455-1258
C 40 69.8 7.4 13511 12 US-10-311-455-254
C 41 69.2 7.3 2126 12 US-10-325-107-59
C 42 69 7.3 17131 12 US-10-311-455-1026
C 43 69 7.3 68611 12 US-10-085-959-182
C 44 68.6 7.2 7057 12 US-10-311-455-1822
C 45 68.6 7.2 7057 12 US-10-240-485-148
```

#### ALIGNMENTS

##### RESULT 1

```
US-10-043-539-1
; Sequence 1, Application US/10043539
; Publication No. US20030114650A1
; GENERAL INFORMATION:
; APPLICANT: Cheung, Ambrose
; APPLICANT: Manna, Adar
; APPLICANT: Zhang, Gongyi
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR AFFECTING VIRULENCE DETERMINANTS IN BACTERIA
; FILE REFERENCE: DC-0199
; CURRENT APPLICATION NUMBER: US/10/043,539
; CURRENT FILING DATE: 2002-01-11
; PRIOR APPLICATION NUMBER: US 60/261,233
; PRIOR FILING DATE: 2001-01-12
; PRIOR APPLICATION NUMBER: US 60/261,607
; PRIOR FILING DATE: 2001-01-12
; PRIOR APPLICATION NUMBER: US 60/289,601
; PRIOR FILING DATE: 2001-05-08
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 947
; TYPE: DNA
; ORGANISM: Staphylococcus aureus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (208)..(552)
; OTHER INFORMATION:
US-10-043-539-1
```

```
Query Match 100.0%; Score 947; DB 14; Length 947;
Best Local Similarity 100.0%; Pred. No. 3.3e-150;
Matches 947; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 1 GTTTCAAAATCGTGAGGTGCATGAAAGTATTGGCATTTCGAAATATAAAAA 60
```

```
Db 1 GTTTCAAAATCGTGAGGTGCATGAAAGTATTGGCATTTCGAAATATAAAAA 60
```



```

; FILE REFERENCE: DC-0199
; CURRENT APPLICATION NUMBER: US/10/043,539
; CURRENT FILING DATE: 2002-01-11
; PRIOR APPLICATION NUMBER: US 60/261,233
; PRIOR FILING DATE: 2001-01-12
; PRIOR APPLICATION NUMBER: US 60/261,607
; PRIOR FILING DATE: 2001-01-12
; PRIOR APPLICATION NUMBER: US 60/289,601
; PRIOR FILING DATE: 2001-03-08
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 19
; LENGTH: 210
; TYPE: DNA
; ORGANISM: Staphylococcus aureus
US-10-043-539-19

Query Match      22.2%; Score 210; DB 14; Length 210;
Best Local Similarity 100.0%; Pred. No. 2e-26;
Matches 210; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTTTTCAAAATCGGTGGAGTGCATGAAAAAGTTATTGGGCATTTTGGAAAAATAAAAA 60
DB 1 GTTTTCAAAATCGGTGGAGTGCATGAAAAAGTTATTGGGCATTTTGGAAAAATAAAAA 60
QY 61 ATATCAATAAGTTGGAGTCATTACCGAATTTTATACCTTTTGTGTAAGTGAACCTTTA 120
DB 61 ATATCAATAAGTTGGAGTCATTACCGAATTTTATACCTTTTGTGTAAGTGAACCTTTA 120
QY 121 TAACATAGTTGGATAGAGTTTTCGATTTAAATACATTTAAATGTGAACCTTGTCAACAAG 180
DB 121 TAACATAGTTGGATAGAGTTTTCGATTTAAATACATTTAAATGTGAACCTTGTCAACAAG 180
QY 181 ATGTGCATCAGAAGGAGTGTTTAATATG 210
DB 181 ATGTGCATCAGAAGGAGTGTTTAATATG 210

RESULT 4
US-10-312-841-1/c
; Sequence 1, Application US/10312841
; Publication No. US20030186277A1
; GENERAL INFORMATION:
; APPLICANT: Epigenomics AG
; TITLE OF INVENTION: Diagnose von bedeutenden genetischen Parametern innerhalb des MHC
; FILE REFERENCE: E01/1208/WO
; CURRENT APPLICATION NUMBER: US/10/312,841
; CURRENT FILING DATE: 2002-12-30
; NUMBER OF SEQ ID NOS: 2
; SEQ ID NO 1
; LENGTH: 3673778
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (3294164)
US-10-312-841-1

Query Match      8.7%; Score 82.4; DB 12; Length 3673778;
Best Local Similarity 48.2%; Pred. No. 0.0012;
Matches 327; Conservative 0; Mismatches 341; Indels 10; Gaps 3;

QY 24 ATGAAAAAGTTATTGGGCATTTTGGAAAAATAAAAAATATCAATAAGTTGGAGTCATTA 83
DB 1714064 ATATAAAATATATAAAAAATATATAAAATATATAAAATATATAAAATATATA 1714005
QY 84 CCGAATTTTATACCTTTTGTGTAAGTGAACCTTTTAAATACATAGTTGGATAGTTTC 143
DB 1714004 TATATAATATAAAATATATATAAAATATATAAAATATATAAAATATATAAAATAT 1713947
QY 144 GATTTAATACATATAAATGTGAACCTTGTCTCAACAAGATGTGTCATCAGAGGAGTGGTTT 203

```

```

DB 1713946 -ATATAAAATATATATAAAATATATAAAATATATAAAATATATAAAATATATAT 1713888
QY 204 AATAATGAGTAAAAATTAUCATATTAATGATTTTAGTCACGCAACATTTCAAGTTAAGAA 263
DB 1713987 AATATATAAAATATATAAAATATATAAAATATATAAAATATATAAAATATATAAAAT 1713828
QY 264 GTTTTTCAGAGATACAAAAAAGAGTTCAATTTGAACCTATGAAGAAATTTTATATTTTAAA 323
DB 1713827 ATATATAAAATATATAAAATATATAAAATATATAAAATATATAAAATATATAAAAT 1713770
QY 324 TCATATTTTAAGAGAGTGAAGTCTAACGCAATCTCATCTAAGAGAGATTCCTAAGTGCCTCAGA 383
DB 1713769 ATATATAAAATATATAAAATATATAAAATATATAAAATATATAAAATATATAAAAT 1713710
QY 384 GTTCAAAACCTTACTATTTTAACTAAAGCTTTTACAAAAGCTTAAAGAGATTTTAAATTTGTTATC 443
DB 1713709 TATAATATATAAAATATATAAAATATATAAAATATATAAAATATATAAAATATATAAAAT 1713650
QY 444 AAGAGAAAAGAGTTTACAGACGAGAGACAGTATTGTTTGTGTACAGATACACAAAA 503
DB 1713649 TATATAAAATATATAAAATATATAAAATATATAAAATATATAAAATATATAAAAT 1713590
QY 504 AGCAAAATATTCAAAAACCTGATTTTCAGATTTAGAGAAATACATTTAAAAATTTAAATCAAGGT 563
DB 1713589 TATATAAAATATATAAAATATATAAAATATATAAAATATATAAAATATATAAAAT 1713535
QY 564 TAATTCGGTTTAAATACATTTGAAGACGATACAAATTTTAAATACGAGAGTTTATTTATTCAGC 623
DB 1713534 TATATAAAATATATAAAATATATAAAATATATAAAATATATAAAATATATAAAAT 1713475
QY 624 ATTGGACATCAAAATTAACCTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAAG 683
DB 1713474 ATATATATAAAATATATAAAATATATAAAATATATAAAATATATAAAATATATAAAAT 1713415
QY 684 ACCAGCCATACCTTTATTT 701
DB 1713414 AATATATAAAATATATAT 1713397

RESULT 5
US-08-781-986A-1599/c
; Sequence 1599, Application US/08781986A
; Publication No. US20030054436A1
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 5255
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage.
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/781,986A
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Benson, Bob
; REGISTRATION NUMBER: 30,446
; REFERENCE/DOCKET NUMBER: PB248PP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504

```





```

QY 102 TTGTTTGAAGTGAACCTTTTATACATAGTTGGATAGAGTTTCGATTTAAATCAATTAATG 161
Db |||||
QY 1530 ATTAATATTTGAATATATTTGATTAATAATATGATGATGATGATGATGATGATGATGAT 1589
Db |||||
QY 162 TGAACCTTGTACACAAAGATGTGCATCAGAGGAGTGTTTAAATGATGATGATGATGATGAT 221
Db |||||
QY 1590 ATAAATATATATATTAATATATAATATATAATATATATATATATATATATATATATAT 1649
Db |||||
QY 222 TGATATATATGATTTAGTCAACGACATTTCAAGTTAAGAGTTTTCAGAGATACAAA 281
Db |||||
QY 1650 AAATATTTGATTAATAATATATAATATATAATATATATATATATATATATATATATAT 1709
Db |||||
QY 282 AAAGAAGTTCAATTTTGAACACTGAAGAAATTTATATTTTAAATCATATTTTAAAGAGTGA 341
Db |||||
QY 1710 ATAAATATTTGATTAATAATATATAATATATAATATATATATATATATATATATATAT 1767
Db |||||
QY 342 GTCTAACGAATCTCATCTCAAGAGATTTGCTAAGTGTGCTCAGAGTTCAAACTTACTATTT 401
Db |||||
QY 1768 ATATATAAATATTTGATTAATAATATATAATATATAATATATATATATATATATATATAT 1827
Db |||||
QY 402 AACTAAAGCTTTACAAAAGCTAAAAGATTTTAAATTTGTTATCAAGAAAGAAAGCTTTTACA 461
Db |||||
QY 1828 AAATATTTGTTTATGCAATATATAATATATAATATATATATATATATATATATATATATA 1887
Db |||||
QY 462 AGACGAAGAGACAGTTATTTGTTATGTTACAGATACACAAAAGCAAAATATTTCAAAACT 521
Db |||||
QY 1888 ATATTATATATTAATATATAATATATAATATATAATATATAATATATAATATATAATATA 1947
Db |||||
QY 522 GATTTTCAGAAATAGAGAAATACATTTAAAAATTTAAATCAAGTTAAATTCGCTTTTAAATACA 581
Db |||||
QY 1948 TATATCATGTTATTTATATATATTTATATATTTTAAATTTGTTTAACTATATATATATAGTT 2007
Db |||||
QY 582 TTGAACGATACAAATTTTAAATCGAAGTTATTTTATTTTACGCAATGGGACATATAAAATTA 641
Db |||||
QY 2008 AACAGAGATTAATAGTTAAATTTAAATATATATATATATATATATATATATATATATATA 2064
Db |||||
QY 642 CTTAAATTTAAATATTTGAAGATGCTTTAAATTTAAAGTTTAAAGTTTAAAGTTTAAAGTTT 701
Db |||||
QY 2065 GTATAAGAGTATTTAGTATTAATGTAATATATAATATATAATATATAATATATAATATATA 2124
Db |||||
QY 702 CAGCTTATTAAGCTTGCACACAGGTACACTAGTCTTTTATTTTAAATATTTTCTTAGAAA 761
Db |||||
QY 2125 TGGTATTTATATTTAAATATATAATATATAATATATATATATATATATATATATATATAT 2184
Db |||||
QY 762 ATCAAGTTTACATCAATATAATTTTTCGCAATATAGCTTTGGATGGTTCCAAAGTATTTT 821
Db |||||
QY 2185 ATATTATAATGATTTATATATTTAAATTTAAATATATAATATATAATATATAATATATAATATA 2244
Db |||||
QY 822 CTCTATAATTTCTGTGCGATTAAGCAAAAATTTCTAACT 858
Db |||||
QY 2245 CAGTTAAATATATATATACAGTATAACATATATAACT 2281
Db |||||

```

# RESULT 8

```

US-10-027-632-113787
; Sequence 113787, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027.632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363

```

```

; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 113787
; LENGTH: 3252
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-113787

Query Match      8.1%; Score 76.6; DB 12; Length 3252;
Best Local Similarity 45.0%; Pred. No. 0.0043;
Matches 368; Conservative 0; Mismatches 444; Indels 5; Gaps 2;

QY 42 ATTTTGAATAAATAAATAATATCAATAAGTTGGAGTCAATACCGAATTTTATACCTTAT 101
Db |||||
QY 102 TTGTTTAGAATGAACCTTTTATACATAGTTGGATAGAGTTTTCGATTTAAATCAATTAATG 161
Db |||||
QY 1530 ATTAATATTTGAATATATTTGATTAATAATATGATGATGATGATGATGATGATGATGAT 1589
Db |||||
QY 162 TGAACCTTGTCTACACAAAGATGTGCATCAGAGGAGTGCTTTAAATCGATGATGATGATGAT 221
Db |||||
QY 1590 ATAAATATTTATATTTAAATATATAATATATAATATATGATTTAAATATATAATATATAT 1649
Db |||||
QY 222 TGNATTAATGATTTAGTCAACGACATTTCAAGTTAAGAGTTTTCAGAGATACAAA 281
Db |||||
QY 1650 AAATATTTGATTTAAATATATAATATATAATATATGATTTAAATATATAATATATATAT 1709
Db |||||
QY 282 AAAGAAGTTCAATTTGCAACTATGAAGAAATTTATATTTTAAATCATATTTTAAAGAGTGA 341
Db |||||
QY 1710 ATAAATTTGATTTAAATATATAATATATAATATATAATATATGTTTAAATATATAATAT 1767
Db |||||
QY 342 GTCTAACGAATCTCATCTCAAGAGATTTGCTAAGTGTGCTCAGAGTTCAAACTTACTATTT 401
Db |||||
QY 1768 ATATAAATAATTTGATTTAAATATATAATATATAATATATGCTTTTAAATATATAATATAT 1827
Db |||||
QY 402 AACTAAAGCTTTACAAAAGCTAAAAGATTTTAAATTTGTTATCAAGAAAGAAAGCTTTTACA 461
Db |||||
QY 1828 AAATATTTGTTTATGCAATATATAATATATAATATATAATATATAATATATAATATATA 1887
Db |||||
QY 462 AGACGAAGAGACAGTTATTTGTTATGTTACAGATACACAAAAGCAAAATATTTCAAAACT 521
Db |||||
QY 1888 ATATTATATATTAATATATAATATATAATATATAATATATAATATATAATATATAATATA 1947
Db |||||
QY 522 GATTTTCAGAAATAGAGAAATACATTTAAAAATTTAAATCAAGTTAAATTCGCTTTTAAATACA 581
Db |||||
QY 1948 TATATCATGTTATTTATATATATTTATATATTTTAAATTTGTTTAACTATATATATATAGTT 2007
Db |||||
QY 582 TTGAACGATACAAATTTTAAATCGAAGTTATTTTATTTTACGCAATGGGACATATAAAATTA 641
Db |||||
QY 2008 AACAGAGATTAATAGTTAAATTTAAATATATAATATATAATATATAATATATAATATATA 2064
Db |||||
QY 642 CTTAAATTTAAATATTTGAAGATGCTTTAAATTTAAAGTTTAAAGTTTAAAGTTTAAAGTTT 701
Db |||||
QY 2065 GTATAAGAGTATTTAGTATTAATGTAATATATAATATATAATATATAATATATAATATATA 2124
Db |||||
QY 702 CAGCTTATTAAGCTTGCACACAGGTACACTAGTCTTTTATTTTAAATATTTTCTTAGAAA 761
Db |||||
QY 2125 TGGTATTTATATTTAAATATATAATATATAATATATAATATATAATATATAATATATAAT 2184
Db |||||
QY 762 ATCAAGTTTACATCAATATAATTTTTCGCAATATAGCTTTGGATGGTTCCAAAGTATTTT 821
Db |||||
QY 2185 ATATTATAATGATTTATATATTTAAATTTAAATATATAATATATAATATATAATATATAATATA 2244
Db |||||
QY 822 CTCTATAATTTCTGTGCGATTAAGCAAAAATTTCTAACT 858
Db |||||
QY 2245 CAGTTAAATATATATATACAGTATAACATATATAACT 2281
Db |||||

```

```

RESULT 9
US-10-027-632-113788
; Sequence 113788, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 113788
; LENGTH: 3252
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-113788

Query Match      8.1%; Score 76.6; DB 12; Length 3252;
Best Local Similarity 45.0%; Pred. No. 0.0013;
Matches 368; Conservative 0; Mismatches 444; Indels 5; Gaps 2;

QY 42 ATTTTTCGAAATGAAATATCAATAGTTGGAGTCATACCGAATTTTATACACTAT 101
DB 1470 ATAATCTTATATTAATTAAGCAATAGCAATATTCATATATTAATAGTATTCATAT 1529

QY 102 TTGTTTGAATGAAATATCAATAGTTGGAGTCATACCGAATTTTATACACTAT 161
DB 1530 ATTAATCTTATATTAATTAAGCAATAGCAATATTCATATATTAATAGTATTCATAT 1589

QY 162 TGAACCTTGCACACAGATGCGATCAGAGGAGTGGTTTAAATGAGTAAATTA 221
DB 1590 ATAAATATTTGATATATTAATTAAGCAATAGCAATATTCATATATTAATAGTATTCATAT 1649

QY 222 TGATATTAATGATTTAGTCAACGCAACATTTCAAGTTAAGAGTTTTCAGAGATACAAA 281
DB 1650 AAAATTTGATATTAATTAAGCAATAGCAATATTCATATATTAATAGTATTCATATAT 1709

QY 342 GTCTAACGAATCTCATCTAAGAGATGCTGAAGTGCACAGTTCCAGCTTCAACCTTACTATTT 401
DB 1768 ATATATAAATATGATATTAATTAAGCAATAGCAATATTCATATATTAATAGTATTCATAT 1827

QY 402 AACTAAAGCTTTACAAAAGCTTAAAGATTTAAATTTGTTATCAAGAAAAAGAGTTTACA 461
DB 1828 AAATATTTGTTTGAATATATATATATAATATTTGTTTAAATATATATATATAATAA 1887

QY 462 AGACGAAGACAGTATTTGTTATGTTACAGATACACAAAAGCAATATTCAAAACACT 521
DB 1888 ATATTATATATTAATATATATATATAATATAATAAATGCAAAATTTATATATAA 1947

QY 522 CATTTTCAGAAATAGAAAGATACATTAATAAATTTAAATCAAGGTTAAATGCGTTTAAATAACA 581
DB 1948 TATATCATGTAATTTATATATTTATATATTTTAAATATGTTTAACTATATATAATAGTT 2007

US-10-027-632-113786
; Sequence 113786, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 113786
; LENGTH: 3252
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-113786

Query Match      8.1%; Score 76.6; DB 13; Length 3252;
Best Local Similarity 45.0%; Pred. No. 0.0013;
Matches 368; Conservative 0; Mismatches 444; Indels 5; Gaps 2;

QY 42 ATTTTTCGAAATGAAATATCAATAGTTGGAGTCATACCGAATTTTATACACTAT 101
DB 1470 ATAATCTTATATTAATTAAGCAATAGCAATATTCATATATTAATAGTATTCATAT 1529

QY 102 TTGTTTGAATGAACTTTTATAACATAGTTGGATAGAGTTTCGATTTTAATACATTAATG 161
DB 1530 ATTAATCTTATATTAATTAAGCAATAGCAATATTCATATATTAATAGTATTCATAT 1589

QY 162 TGAACCTTGCACACAGATGCGATCAGAGGAGTGGTTTAAATGAGTAAATTA 221
DB 1590 ATAAATATTTGATATATTAATTAAGCAATAGCAATATTCATATATTAATAGTATTCATAT 1649

QY 222 TGATATTAATGATTTAGTCAACGCAACATTTCAAGTTAAGAGTTTTCAGAGATACAAA 281
DB 1650 AAAATTTGATATTAATTAAGCAATAGCAATATTCATATATTAATAGTATTCATATAT 1709

QY 342 GTCTAACGAATCTCATCTAAGAGATGCTGAAGTGCACAGTTCCAGCTTCAACCTTACTATTT 401
DB 1768 ATATATAAATATGATATTAATTAAGCAATAGCAATATTCATATATTAATAGTATTCATAT 1827

QY 402 AACTAAAGCTTTACAAAAGCTTAAAGATTTAAATTTGTTATCAAGAAAAAGAGTTTACA 461
DB 1828 AAATATTTGTTTGAATATATATATATAATATTTGTTTAAATATATATATATAATAA 1887

QY 462 AGACGAAGACAGTATTTGTTATGTTACAGATACACAAAAGCAATATTCAAAACACT 521
DB 1888 ATATTATATATTAATATATATATATAATATAATAAATGCAAAATTTATATATAA 1947

QY 522 CATTTTCAGAAATAGAAAGATACATTAATAAATTTAAATCAAGGTTAAATGCGTTTAAATAACA 581
DB 1948 TATATCATGTAATTTATATATTTATATATTTTAAATATGTTTAACTATATATAATAGTT 2007

```



```

; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 113788
; LENGTH: 3252
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-113788

Query Match      8.1%; Score 76.6; DB 13; Length 3252;
Best Local Similarity 45.0%; Pred. No. 0.0013;
Matches 368; Conservative 0; Mismatches 444; Indels 5; Gaps 2;

QY 42 ATTTTTCGAAATGAAAAATATCAATAAGTTGGAGTCATTACCGAATTTTATACCTAT 101
DB 1470 ATAATTCCTTATATATATATTAAGATAGCAATATATCATATATTAATAGTATTCATAT 1529
QY 102 TGTGTTAGAATGAACCTTTATACATAGTTGGATAGAGTTTTCGATTTTAAATACATTAATG 161
DB 1530 ATTAATATTTATGAATATATTCATATATAATATATGATGATACAAATCAATATATATATAT 1589
QY 162 TGAACCTTGTACACACAGAGTGTGCATCAGACGAGTGGTTTATATATGATGTAATTA 221
DB 1590 ATAATAATATATATTAATAATATATATATATATAATATTTGTTATATTAATAATATATATAT 1649
QY 222 TGATATTAATGATTTAGTCAACGCAACATTTCAAGTTTAAAGAGTTTTCAGAGATACAAA 281
DB 1650 AAATATTTGATATTAATAATATATATAATAATAATAATAATTTGTTATATTAATAATATATATAT 1709
QY 282 AAAGAGTTTCAATTTGAACATGATGAAGAAATTTATATTTTAAATCATATTTTTRAGAAGTGA 341
DB 1710 ATAAATATTTGATATTAATAATATATATAATATAATAATATTTGT--ATATTAATAATATATA 1767
QY 342 GTCTAAACGAAATCTCATCTAAAGAGATTTGCTAAGTGTCTCAGAGTTCAACCTTACTATTT 401
DB 1768 ATATATAAATATTTGATATTAATAATATATATATATATATTTGTTTATTAATAATATATATAT 1827
QY 402 AACTAAAGCTTTACAAAAGCTTAAAGATTTTAAATTTGTTATCAAGAAAAGAGTTTACAA 461
DB 1828 AAATATTTGTTTATTGCAATATATATAATAATAATATTTGTTTATTAATAATATATAATATA 1887
QY 462 AGACCAAGAGACAGTTATTTGTTTATGTTACAGATACACAAAAGCAAAATATTCAAAACCT 521
DB 1888 ATATATATATTAATAATATTAATAATATATATATAATAATAATAATGCAAAATTTATATATA 1947
QY 522 GATTTCAAGAAATAGAGAATACATATAATAATAATAATAATAATAATAATAATAATAATAATA 581
DB 1948 TATATCATGTTATTTATATATATTTATATATATTTTAAATGTTTAACTATATATAATAAGTT 2007
QY 582 TTGACAGTACATTTTATTAATACGAGTATTTTATTCAGCATTTGGGACATTAATAATTA 641
DB 2008 AACAGATTAATAAGTTAAATTTATATATTAATAATTAATTTTAAATTTTAACTAAATA 2064
QY 642 CTTAAAAATTTAAATATTGAAGATCTTTAATTTAAAGTTAAAGCCACCATACCTTATTT 701
DB 2065 GTATAAGATTTATAGTATTAATGTAATATATTAATAATTAATAATTTTAAATATAGTTATTA 2124
QY 702 CAGCTTATTAAGCTTGACACAGGTACACTAGCTTTTATTTTATTAATTTTCTTGTAGAA 761
DB 2125 TGGTATTATATATTAAATATATATAATATGAATATGTTATGATATATATTTGTTATTAATTT 2184
```

```

QY 762 ATCAAGTTTACGATCATATAAATATTTTCTCGGATATAGCTTTGGATGGTTCCAAAGTATTTT 821
DB 2185 ATATTATAATGTATTATATATTTAAATTTAAATATATTAACTTATTCATATGTTACCATACA 2244
QY 822 CTCATAAATTTGTGCGGATAGCAAGCAAAAATTTCTAACT 858
DB 2245 CAGTTAAATATATATTACAGTATAAACAATATTAACT 2281

RESULT 13
US-09-912-020-468/c
; Sequence 468, Application US/09912020
; Patent No. US20020045592A1
; GENERAL INFORMATION:
; APPLICANT: Zyskind, Judith
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Trawick, John
; APPLICANT: Forsyth, R. Allyn
; APPLICANT: Proelich, Jamie M.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: GENES IDENTIFIED AS REQUIRED FOR PROLIFERATION IN
; TITLE OF INVENTION: ESCHERICHIA COLI
; FILE REFERENCE: ELITRA.001DV1
; CURRENT APPLICATION NUMBER: US/09/912,020
; PRIOR FILING DATE: 2001-07-23
; PRIOR APPLICATION NUMBER: 09/492,709
; PRIOR FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: 60/117,405
; PRIOR FILING DATE: 1999-01-27
; NUMBER OF SEQ ID NOS: 485
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 468
; LENGTH: 529
; TYPE: DNA
; ORGANISM: Escherichia coli
US-09-912-020-468

Query Match      8.0%; Score 75.6; DB 9; Length 529;
Best Local Similarity 76.2%; Pred. No. 0.0011;
Matches 93; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

QY 1 GTTTTCAAAAATCGGTGGAGTGCATGAAAAAGTTATTTGGCATTTTTTGGCAATTTTTTGAATAA 60
DB 377 GTTTCAAAATCGGTGGAGTGCATGACAAAGTCATCGGCATTATCTGGAACATAAAACA 318
QY 61 ATATCAATAAGTTGGAGTCATACCGAATTTTATATCTATTGTTTGAATGAACCTTTA 120
DB 317 CTATCAATAAGTTGGAGTCATACCGACATTAAAGTAACCTTTTTTGAATTAGAGATAA 258
QY 121 TA 122
DB 257 TA 256

RESULT 14
US-10-122-085A-5/c
; Sequence 5, Application US/10122085A
; Publication No. US20030172404A1
; GENERAL INFORMATION:
; APPLICANT: Greenlee, Winner and Sullivan, P.C.
; APPLICANT: JOHN, Peter C.L.
; APPLICANT: ZHANG, Kerong
; APPLICANT: SEK, Francis J.
; APPLICANT: VAN CAMP, Wim J.
; TITLE OF INVENTION: METHOD OF MODIFYING PLANT CHARACTERS BY THE TARGETED EXPRESSION OF
; FILE REFERENCE: Docket 19-00A
; CURRENT APPLICATION NUMBER: US/10/122,085A
; PRIOR FILING DATE: 2002-04-10
; PRIOR APPLICATION NUMBER: US 09/513,504
; PRIOR FILING DATE: 2000-02-25
```

; PRIOR APPLICATION NUMBER: US 60/121,870  
; PRIOR FILING DATE: 1999-02-26  
; PRIOR APPLICATION NUMBER: US 60/149,049  
; PRIOR FILING DATE: 1999-08-16  
; NUMBER OF SEQ ID NOS: 5  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 5  
; LENGTH: 11777  
; TYPE: DNA  
; ORGANISM: Artificial  
; FEATURE:  
; OTHER INFORMATION: Cloning vector pBin19  
US-10-122-085A-5

Query Match 7.9%; Score 75.2; DB 12; Length 11777;  
Best Local Similarity 86.5%; Pred. No. 0.0033;  
Matches 83; Conservative 0; Mismatches 13; Indels 0; Gaps 0;  
QY 1 GTTTCAAAATCGGTGGAGTGCATGAAAAGTTATTGGGCATTTTGGAAAATAAAAA 60  
DB 1401 GTTCTCAAAATCGGTGGAGTGCATGACAAAGTCATCGGCATTTCTGAACATAAAACA 1342  
QY 61 ATATCAATAAGTTGGAGTCATTACCGAATTTTATA 96  
DB 1341 CTATCAATAAGTTGGAGTCATTACCCCAATTATGATA 1306

RESULT 15  
US-10-033-190-5  
; Sequence 5, Application US/10033190  
; Publication No. US20020133848A1  
; GENERAL INFORMATION:  
; APPLICANT: Exelixis Plant Sciences, Inc.  
; TITLE OF INVENTION: IDENTIFICATION AND CHARACTERIZATION OF AN ANTHOCYANIN MUTANT (ANT  
; FILE REFERENCE: EP01-002C  
; CURRENT APPLICATION NUMBER: US/10/033,190  
; CURRENT FILING DATE: 2001-10-29  
; PRIOR APPLICATION NUMBER: US 60/244,685  
; PRIOR FILING DATE: 2000-10-30  
; NUMBER OF SEQ ID NOS: 5  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 5  
; LENGTH: 12241  
; TYPE: DNA  
; ORGANISM: PAG2370  
US-10-033-190-5

Query Match 7.9%; Score 75.2; DB 13; Length 12241;  
Best Local Similarity 86.5%; Pred. No. 0.0034;  
Matches 83; Conservative 0; Mismatches 13; Indels 0; Gaps 0;  
QY 1 GTTTCAAAATCGGTGGAGTGCATGAAAAGTTATTGGGCATTTTGGAAAATAAAAA 60  
DB 10841 GTTCTCAAAATCGGTGGAGTGCATGACAAAGTCATCGGCATTTCTGAACATAAAACA 10900  
QY 61 ATATCAATAAGTTGGAGTCATTACCGAATTTTATA 96  
DB 10901 CTATCAATAAGTTGGAGTCATTACCCCAATTATGATA 10936

Search completed: November 17, 2003, 12:42:25  
Job time : 374 secs

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 17, 2003, 09:13:37 ; Search time 2414 Seconds  
(without alignments)  
9534.523 Million cell updates/sec

Title: US-10-043-539A-1

Perfect score: 947

Sequence: 1 gtttcaaatcgttgagg.....gtttcataaactgaatcgat 947

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*

1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estmu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_htc:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_htc:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: em\_gss\_hum:\*  
18: em\_gss\_inv:\*  
19: em\_gss\_pln:\*  
20: em\_gss\_vrt:\*  
21: em\_gss\_fun:\*  
22: em\_gss\_mam:\*  
23: em\_gss\_mus:\*  
24: em\_gss\_pro:\*  
25: em\_gss\_rtd:\*  
26: em\_gss\_phg:\*  
27: em\_gss\_vrl:\*  
28: gb\_gss1:\*  
29: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	115	12.1	1101	29	CNS00EVL
2	108.8	11.5	1200	13	AL069706 Drosophila
3	100.8	10.6	1200	13	BX415878 BX415878
4	99.8	10.5	1056	13	BX415058 BX415058

C	5	99	10.5	1101	29	CNS0039G
C	6	98.8	10.4	1201	13	BX439779
C	7	97.8	10.3	1200	13	BX437739
C	8	97.4	10.3	1045	13	BX456814
C	9	96.8	10.2	1200	13	BX437758
C	10	93.8	9.9	1201	13	BX462207
C	11	93.2	9.8	1201	13	BX420717
C	12	92.8	9.8	1201	9	AL536104
C	13	92.2	9.7	887	10	BF274559
C	14	92.2	9.7	1123	14	CD325190
C	15	91.8	9.7	1200	13	BX437758
C	16	91.4	9.7	1201	9	AL565455
C	17	88.4	9.3	1200	13	BX436510
C	18	86.4	9.1	1101	29	CNS0021J
C	19	86.2	9.1	1201	13	BX361152
C	20	86.2	9.1	1201	13	BX458623
C	21	86	9.1	1101	29	CNS00EO7
C	22	85.8	9.1	945	29	CNS04D0K
C	23	85.6	9.0	1201	13	BX443774
C	24	85.2	9.0	1101	29	CNS00EO7
C	25	85	9.0	1098	13	BX377526
C	26	84.8	9.0	1201	13	BX335216
C	27	84.6	8.9	1201	13	BX340226
C	28	84.4	8.9	1201	9	AL536104
C	29	84.2	8.9	1201	13	BX461824
C	30	83.8	8.8	1128	13	BX337294
C	31	83	8.8	1061	13	BX437039
C	32	82.8	8.7	836	29	CNS01100
C	33	82.2	8.7	1101	29	CNS00EPO
C	34	82	8.7	1101	29	CNS000B8
C	35	81.6	8.6	1201	13	BX366070
C	36	81.6	8.6	1201	13	BX446437
C	37	81.4	8.6	1201	29	CNS0167M
C	38	81.2	8.6	1101	29	CNS0039R
C	39	81.2	8.6	1101	29	CNS00LT2
C	40	81.2	8.6	1201	13	BX406288
C	41	81	8.6	939	14	BY720774
C	42	80.8	8.5	867	29	CNS075BG
C	43	80.8	8.5	1201	13	BX334085
C	44	80.6	8.5	781	29	CNS009DO
C	45	80.6	8.5	1201	13	BX357882

#### ALIGNMENTS

RESULT 1  
CNS00EVL 1101 bp DNA linear GSS 04-JUN-1999  
Drosophila melanogaster genome survey sequence T7 end of BAC:  
BACR29B23 of RFLI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.  
AL069706 GI:4949849  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
Drosophila melanogaster (fruit fly)  
Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
Genoscope.  
Direct Submission  
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)  
Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammoss in Pieter de Jong's laboratory in the Department of







```

/tissue type="THYMUS"
/clone lib="Homo sapiens THYMUS"
/Note="Vector: pCMVSPORT 6; 1st strand cDNA was primed
with a NotI-oligo(dT) primer. Five prime end enriched,
double-strand cDNA was digested with Not I and cloned into
the Not I and EcoRV sites of the pCMVSPORT 6 vector.
Library was not normalized."
BASE COUNT 454 a 53 c 56 g 318 t 175 others
ORIGIN
Query Match 10.5%; Score 99.8; DB 13; Length 1056;
Best Local Similarity 36.3%; Pred. No. 0.00042; Indels 6; Gaps 1;
Matches 279; Conservative 106; Mismatches 378;
QY 27 AAAAAAGTATTGGGCAATTTTGGAAATAAAAAATCAATAAGTTGGAGTCATTACCG 86
Db 243 AAATAGTAATAAAAAATTTACAAAATAATGTAATGTAATAAAGTGAAGCTGAGACTACATAT 302
QY 87 AATTTTACTTACTTTGTTAGTAGAACCTTTTATACATAGTTGGATAGATTTTCGAT 146
Db 303 TCTAGTATTAATAAGTCAAAACAACAATAGTTTATAAACTTGGTCTGCTGAATATTTT 362
QY 147 TTAATACATTAATGTAACCTTGTCAACAAGATGTCATCAGAAGGAGTGGTTTAAT 206
Db 363 ATATATATATTTTATATATATATATAWAAAAAATAAAAAAATAAAAAAATATATAT 422
QY 207 AATGAG-----TAAATTAATGATTAATGATTTAGTCAACGCACATTTCAAGTTAA 260
Db 423 AWTTKTTTTTTTWTATTTATAAAAAARWATATATTTAAAAAATAAAAAAATTTTWA 482
QY 261 GAAGTTTTTCAGAGATACAAAAAGAGTTCATTTGAACATCATCAAGAAATTTATATTTT 320
Db 483 AAWTTTTTATWAAATTTWAAAAAATAAAATTTWKKKWTATATAAAAAAATAAAAAA 542
QY 321 AAATCATATTTTAAGAGAGTGTCTAAGCAATCTCATCTAAGAGATTTGCTAAGTCTC 380
Db 543 AAAAATATATATATTTATATATATAWAAAAAATAAAAAAATAAAAAAATAAAAAA 602
QY 381 AGAGTTCAACCTTACTATTTTAACTAAGCTTTTCAAAAGCTAAAGATTTAAATTTGT 440
Db 603 TTTGAAWATGTTTTTTTCTTTTWTWATWATTTTWTATADATRTKTTTTTWTATDW 662
QY 441 ATCAAGAAAAAGAGTTTACAGACGAAAGAACAGTATTGTTATGTTACAGATACACA 500
Db 663 DATWAAATATWTBTGAGWAAAAAATAKTKATKTKTAWAAAAAATAAAAAAATAAAA 722
QY 501 AAAGCAATATTTCAAAACATGTTTCAGAAATAGAGATACATTAATAAATTAATCAA 560
Db 723 AAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATA 782
QY 561 GGTAAATGCGTTTAAATAACATTTGAACGATAAACAATTTTATTAACGAAGTTATTTATTC 620
Db 783 TTTTITTTTWWDDDDDDDDTTTITTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 842
QY 621 AGCATTGGGACATAAAATACTAAATAATTAATATATGAAGTCTTTTAAATTAAGTTA 680
Db 843 WWWWWWWAAAAAATAAAATTTTAAATTTTATAAATAAATAAATTTTTRWAAAAAATA 902
QY 681 AAGACCAACCCATCTTATTTTCAAGTTATTAAGCTTCGACACAGGTCACATGCTCTTTT 740
Db 903 AAAAAATTAATAAATAAATTTWAAAAAATAAATAAATAAATAAATAAATAAATAAATA 962
QY 741 ATTTTATATTTCTTGAATAATCAAGTTTACGATCATATAATTTCT 789
Db 963 TAGGAAGGGTTTAAAAAATAATGATTAATATGTTKTTTTTAAAAAATTTAT 1011

```

RESULT 5

CNS0039G/c

LOCUS

DEFINITION

Drosophila melanogaster genome survey sequence TBT3 end of BAC # BACR08K10 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.

ACCESSION  
AL063921.1 GI:4941778  
VERSION  
GSS  
KEYWORDS  
Drosophila melanogaster (fruit fly)  
SOURCE  
Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
Ephydroidea; Drosophilidae; Drosophila.  
1 (bases 1 to 1101)  
Genoscope.  
Direct Submission  
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :  
BP 191 91006 EVRY cedex - FRANCE (E-mail : segref@genoscope.cns.fr  
- Web : www.genoscope.cns.fr)

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Determination of this BAC-end sequence was carried out as part of a  
collaboration with the Berkeley Drosophila Genome Project (BDGP).  
The BDGP is constructing a physical map of the Drosophila  
melanogaster genome using these BACs. For further information  
please see <http://www.fruitfly.org> the BDGP Drosophila  
melanogaster BAC library was prepared by Kazuo Osoegawa and  
Aaron Mammeter in Pieter de Jong's laboratory in the Department of  
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,  
NY. The library is named RPCI-98 and was constructed by partial  
EcoRI digestion of Drosophila DNA provided by the BDGP from the  
isogenic strain v2; cn bw sp, the same strain used for the BDGP's  
P1 and EST libraries. A more detailed description of the library  
and how to order individual BAC clones, the entire library, or  
filters for hybridization from the BACPAC Resource Center can be  
found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).

FEATURES

Location/Qualifiers

source

1..1101  
/organism="Drosophila melanogaster"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:7227"  
/clone="BACR08K10"  
/clone\_lib="RPCI-98"  
/note="end : TBT3"

BASE COUNT 201 a 64 c 131 g 202 t 503 others  
ORIGIN

Query Match

Best Local Similarity 10.5%; Score 99; DB 29; Length 1101;

Matches 132; Conservative 303; Mismatches 265; Indels 1; Gaps 1;

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

```
Db 678 TWDRTDAMWADAATTTTDDTDDKDRRRKRGARRRRRTTARAANDWNTWKANDWAKW 619
QY 555 AATCAGGTTAATTCGGTTTAAATACATGTAACAGTAACAATTTATTAATACGAGTTAT 614
Db 618 DWKTEADRWDRWAADTWDARXADRWAKARAWRARRDRRAARDRRWTTKGKTTTATW 559
QY 615 TTATTGAGCATGGGACACATAAAATTAACCTTAAATTTAAATTTGAAGATGCTTTAATTA 674
Db 558 TTWAAABAAWAAWAAWATTTATWTTTWTWTTTWTWTTTWTWTTTAAWAAWTAWAA 499
QY 675 AAGTTAAACACAGCCATACCTTATTGAGCTTAAAGCTTGACACAGGTACACTAGT 734
Db 498 WTAABAAWAAWAAWAAATTTTTTTTTTTTAAWAAWTAWTTTWTWTTTAAWAAATTTT 439
QY 735 CTTTATTATTAAATTTTCTTTAGAAAATCAAGTTTACGAT 775
Db 438 TWTWTTWAAATATTTTTWTWTTTAAWAAWAAWAAWTTWTTAT 398

RESULT 6
LOCUS BX439779 Homo sapiens PLACENTA Homo sapiens cDNA clone CS0DE014YF05
DEFINITION 3-PRIME, mRNA sequence.
ACCESSION BX439779
VERSION BX439779.1 GI:30771778
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Li.W.B., Gruber,C., Jessee,J. and Polayes,D.
Full-length cDNA libraries and normalization
Unpublished
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 3370.r For
more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CS0DE014CC03NP1&cluster=3370.r. Contact :
Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Paraday Avenue Genoscope sequence ID : CS0DE014CC03NP1.

FEATURES
Source
1. 1201
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DE014YF05"
/tissue_type="PLACENTA"
/clone_lib="Homo sapiens PLACENTA"
/note="Vector: pCMVSPORT 6; 1st strand cDNA was primed
with a NotI-oligo(dT) primer. Five prime end enriched
double-strand cDNA was digested with Not I and cloned into
the Not I and EcoRV sites of the pCMVSPORT 6 vector.
Library was not normalized."
BASE COUNT 455 a 150 c 90 g 299 t 207 others
ORIGIN
Query Match 10.4%; Score 98.8; DB 13; Length 1201;
Best Local Similarity 36.0%; Pred. No. 0.00055;
Matches 207; Conservative 94; Mismatches 274; Indels 0; Gaps 0;

QY 45 TTTTGAAAATAAAAAATATCAATAAGTTGGAGTCATTACCGAATTTTATACCTATTG 104
Db 614 HTGYTAAXATATATTATWTKRGRTTTAAATATATWTTTAAATWTTGTTGATATWTKTT 673
QY 105 TTTAGATGAACCTTTATACATAGTTGGATAGAGTTTCGATTATACATTAATATGTA 164
```

```
Db 674 TTWAAATAATWTTTTTTTATTATTAATTTTAAWAAWAAATTTATTAATAATWAAWTAA 733
QY 165 ACCTTGCATCAACAAGATGCGCATCAGAGGAGTGGTTTAAATAGTAAATTAATGA 224
Db 734 AATTWTAHAATAAAAAATWTAATTAATAATAAATKTAATTTWAAAATTTAAATTTT 793
QY 225 TATTAAATGATTAGTCAAGCGCAACATTCAGGTTAAGAAAGTTTTCAGAGATACAAAAA 284
Db 794 AATATAATTTAAWAAATTTTATAATATWTTTWTATATAATWTTTAAWTTTAAAWTA 853
QY 285 GAAGTTCAATTCGAATCATTGAAGAAATTTATATTTTAAATCATATTTTAAAGAGTGAGTC 344
Db 854 AAWWAAWAAAAWAAWTTTAAWAAWTTTAAWAAWTTTAAWTTTAAWAAWAAATTTT 913
QY 345 TAAAGAAATCTCATCTAAAGAGATTCGCTAAGTCTCAGAGTTCCAAACCTTACTATTAAAC 404
Db 914 TAWWATAWTAATWTAATWTAATWTTTAAWTTTAAWTTTAAWTTTAAWTTTAAWTTTAAW 973
QY 405 TAAAGCTTTTACAAAAGCTAAAGATTTAAATTTGTTTATCAAGAAAGAAAGTTTACAAGA 464
Db 974 TTAATAATWTAATTAATAAAAAWAAWAAWAAWAAWAAWTTTAAWTTTAAWTTTAAW 1033
QY 465 CGAAAGACAGATTATGTTTATGTTTATGTTTACAGATACACAAAGGCAAAATTTCAAAACCTGAT 524
Db 1034 TATTWAAAWWAAWTTTWTWTTTWTWTTTAAWTTTAAWTTTAAWTTTAAWTTTAAW 1093
QY 525 TTCGAATTTAGAAAGATAATACATTAAAAATTTAAATCAAGGTTAAATTCGTTTAAATCAATG 584
Db 1094 WAAATAATTTAAWAAWAAWAAWAAWAAWAAWAAWAAWAAWAAWAAWTTTAAWTTTAAW 1153
QY 585 AAGATAACAAATTTATTAATACGAAGTTTATTATT 619
Db 1154 TAAWAAATTAATAATTAATAWAAWAAWAAWAAWAAWAAWAAWAAWAAWAAWAAW 1188

RESULT 7
LOCUS BX437739/c Homo sapiens THYMUS Homo sapiens cDNA clone CS0CAP008YG24
DEFINITION 3-PRIME, mRNA sequence.
ACCESSION BX437739
VERSION BX437739.1 GI:31018315
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Li.W.B., Gruber,C., Jessee,J. and Polayes,D.
Full-length cDNA libraries and normalization
Unpublished
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 534.r For
more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CS0CAP008BD12NP1&cluster=534.r. Contact :
Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Paraday Avenue Genoscope sequence ID : CS0CAP008BD12NP1.

FEATURES
Source
1. 1200
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0CAP008YG24"
/tissue_type="THYMUS"
/clone_lib="Homo sapiens THYMUS"
/note="Vector: pCMVSPORT_6; 1st strand cDNA was primed
```



Qy	843	AGCAAAATTTCTAACTCAAAACCATCTGTAGGCAATTTGAGAATACGACACGACATC	907
Db	442	TTTAAATTTAAATTTTTTYYAAATTTATTTAAAWYAAAAAATAAAAAAAAAAAAAA	383
Qy	903	GGATGATTGCTATAAGAACTAATGGTTTCATAAACTGAA	942
Db	382	WTATTTYYAATTATTTYYAAAVATTTYYAATTTTTTAA	343
RESULT 9			
LOCUS	BX437758	1200 bp	linear
DEFINITION	BX437758 Homo sapiens THYMUS Homo sapiens cDNA clone CS0CAP008YB01		EST 15-MAY-2003
VERSION	5-PRIME, mRNA sequence.		
KEYWORDS	BX437758		
SOURCE	BX437758.1 GI:30773605		
ORGANISM	Homo sapiens (human)		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
AUTHORS	Li, W.B., Gruber, C., Jesse, J. and Polayes, D.		
TITLE	(bases 1 to 1200)		
JOURNAL	Full-length cDNA libraries and normalization		
COMMENT	Unpublished		
	Contact: Genoscope		
	Genoscope - Centre National de Sequencage		
	BP 191 91006 EVRY cedex - France		
	Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr		
	Library was constructed by Life Technologies, a division of		
	Invitrogen. Contact : Feng Liang Email : fliang@lifetech.com URL :		
	http://fulllength.invitrogen.com/		
	Invitrogen Corporation 1600		
	Faraday Avenue Genoscope sequence ID : CS0CAP008CA01Q01.		
FEATURES	Location/Qualifiers		
source	1..1200		
	/organism="Homo sapiens"		
	/mol_type="mRNA"		
	/db_xref="taxon:9606"		
	/clone="CS0CAP008YB01"		
	/tissue_type="THYMUS"		
	/clone_lib="Homo sapiens THYMUS"		
	/note="Vector: pCMVSPORT 6; 1st strand cDNA was primed		
	with a NotI-oligo(dT) primer. Five prime end enriched,		
	double-strand cDNA was digested with Not I and cloned into		
	the Not I and EcoRV sites of the pCMVSPORT 6 vector.		
	Library was not normalized."		
BASE COUNT	515 a 30 c 71 g 310 t 274 others		
ORIGIN			
Query Match	10.2%; Score 96.8; DB 13; Length 1200;		
Best Local Similarity	30.4%; Pred No. 0,001;		
Matches	235; Conservative 170; Mismatches 365; Indels 4; Gaps 1		
Qy	50	AAATATAAAATATCAATAGTTGGATGATACCGAATTTTATACATTATTGTTTAG	109
Db	264	AAAAAATAAAAWAAATATTTTAAARATWDGATKADDAATTTGGGGDWT	323
Qy	110	AATGAACCTTTAATCATAGTTGGATAGAGTTTCGATTTAATACATTAATGTCGAACCTT	169
Db	324	AARRDKTTTTTTTNAAAARATAXKGGRGADATWTTTATTTATTTTATTTTATTTT	383
Qy	170	GCTACAAAGATGTCACAGAGGATGGTTTAAATAGTAGTAAATTAATGATATTA	229
Db	384	AKAWADAGAAATTTTTTTTWRAGRGARKRDGTTDTRAADWWTTTAAAWWWDKKTTT	443
Qy	230	ATGATTTAGTCAACCAACATTTCAAGTT---AAGAAGTTTTTTCAGAGATACAAAAAG	285
Db	444	TTTTKGRRAAAATTTTTTTTTTTTTTTTWTWAKAAAAAARATTTTWRAAAAAATTTT	503
Qy	286	AAGTTCATTTGAACATGAAGAAATTTATATTTTAAATCAATTTTAAAGAGTGAGTCT	345
Db	504	ARGRARGATTTTTTTTAAWWCGGAGRTAAWTTTWTTTTTTTTTTTTTTTTAAWTTATA	563

the Not I and EcoRV sites of the pcwvSPORT 6 vector.  
Library was not normalized."

```

BASE COUNT 308 a 141 c 153 g 522 t 77 others
ORIGIN
Query Match 9.9%; Score 93.8; DB 13; Length 1201;
Best Local Similarity 40.9%; Pred. No. 0.0025;
Matches 269; Conservative 48; Mismatches 341; Indels 0; Gaps 0;
QY 173 ACAACAGATGTCATCAGACGAGTGGTTTAAATAGTAAATTAATGATTAATG 232
Db 1087 AAAAAAACAACAAAAAATAAATAATTTTAAATAATTAATTAATTAATTAAT 1028
QY 233 ATTATGCTCAACCAACATTCAGTTTAAAGTAAAGTAAAGTAAAGTAAAGT 292
Db 1027 TTTAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 968
QY 293 ATTGCACTATGACGAAATTTTATTTTAAATCATATTTTAAAGAGTGAAGTCA 352
Db 967 AWATTTTAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 908
QY 353 TCTCATCTAAAGAGATGCTAGTCTCAGATTCAGATTCAGATTCAGATTCAG 412
Db 907 TTTAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 848
QY 413 TACAAAGCTAAAGATTTAAATTTGTTTATCAAGAAAGAGTAAAGTAAAGT 472
Db 847 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 788
QY 473 CAGTTATGTTTATGTTACAGATACACAAAGCAATTAATCAAAAGTAAAT 532
Db 787 TTTTAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 728
QY 533 TAGAAGATACATTAATAAATAAATAAATAAATAAATAAATAAATAAATAA 592
Db 727 AAATTTTAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 668
QY 593 CAATTTATTAACGAGATTTATTCAGATTCAGATTCAGATTCAGATTCAG 652
Db 667 AAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 608
QY 653 ATATTGAGATGCTTTAATAAGTTAAAGTAAAGTAAAGTAAAGTAAAGTAA 712
Db 607 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 548
QY 713 GCTTACACAAAGGTACACTAGTCTTTTATTTTAAATTTTCTTAAAGTAA 772
Db 547 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 488
QY 773 GATCAATAATTTTCTCGGATAGCTTTCGATGCTTTCGATGCTTTCGAT 830
Db 487 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 430

```

```

RESULT 11
LOCUS BX420717
DEFINITION BX420717 Homo sapiens FETAL BRAIN Homo sapiens cDNA clone
ACCESSION CS0DF033YC10 5-PRIME, mRNA sequence.
VERSION BX420717
KEYWORDS BX420717.1 GI:30658996
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1. (bases 1 to 1201)
AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished
CONTACT: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France

```

Email: segrf@genoscope.cns.fr, Web : www.genoscope.cns.fr  
Library was constructed by Life Technologies, a division of  
Invitrogen. Contact : Feng Liang Email : fliang@lifetech.com URL :  
http://fulllength.invitrogen.com/Invitrogen Corporation 1600  
Paradise Avenue Genoscope sequence ID : CS0DF033BB05Qp1.

#### FEATURES

```

Location/Qualifiers
source 1..1201
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DF033YC10"
/tissue_type="FETAL BRAIN"
/dev_stage="fetal"
/clone_lib="Homo sapiens FETAL BRAIN"
/notes="Organ: Brain; Vector: pcwvSPORT_6; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-strand cDNA was digested with Not I and
cloned into the Not I and EcoRV sites of the pcwvSPORT 6
vector. Library was not normalized."
BASE COUNT 585 a 115 c 108 g 299 t 94 others
ORIGIN

```

```

Query Match 9.8%; Score 93.2; DB 13; Length 1201;
Best Local Similarity 42.4%; Pred. No. 0.003;
Matches 271; Conservative 52; Mismatches 311; Indels 5; Gaps 2;
QY 50 AAAAAAATAAATCAATAAGTTGAGTCAATACCGATTTTATCTATTTCTTAG 109
Db 521 AAAAAAATAAATCAATAAGTTGAGTCAATACCGATTTTATCTATTTCTTAG 580
QY 110 AATGAACCTTTATAACATAGTTGGATAGAGTTTTCGATTTTAAATCAAT 165
Db 581 TATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 640
QY 166 CTTGCTACACAGATGTCATCAGAGGAGTGGTTTAAATCAATCAATCAAT 225
Db 641 AATTTTAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 700
QY 226 AATAATGATTTAGTCAACGCAACATTTCAAGTTTAAAGTAAAGTAAAG 285
Db 701 WAAATTTTATTTAAATAAATAAATAAATAAATAAATAAATAAATAAATA 760
QY 286 AAGTCAATTTGAACTATGAGAAATTTATTTTAAATCAATTTTAAAGAGT 345
Db 761 WAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 820
QY 346 AACGAAATCTCATCTAAAGAGATTCGTAAGTCTCAGAGTTCACACCTT 405
Db 821 AAATTTATTTAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 880
QY 406 AAGCTTTTCAAAAGCTAAAGATTTAAATTTGTTTATCAAAAGAAAGAGT 465
Db 881 AAAAAATWAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 940
QY 466 GAAAGACAGTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTAT 525
Db 941 ANATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1000
QY 526 TCAGAAATAGAGAAATACATTTAAATTTAAATTTAAATTTAAATTTAAAT 585
Db 1001 TATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 1059
QY 586 AGATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 645
Db 1060 TAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 1119
QY 646 AATTTTAAATTTGAGATGCTTTTAAATTTAAAGTAAAGTAAAGTAAAG 684
Db 1120 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 1158

```

RESULT 12  
AL536104/c



QY 381 AGAGTTCAACCTTACTATTAACTAAAGCTTTTACAAAGCTAAAGATTTTAAATTTGTT 440  
 Db 479 NAAAAATATTAATTAATAATTAATAATTAATAATTAATAATTAATAATTAATAATTAAT 538  
 QY 441 ATCAAGAGAAAGAGTTTACAGAGCAAGAGACAGTTTATTTTATGTTTACAGATPACACA 500  
 Db 539 AACTAAAAATTAATAATTAATAATTAATAATTAATAATTAATAATTAATAATTAATAAT 598  
 QY 501 AAGAGCAATATTCAGAACTGATTTTCAAGATTAAGAGATACATTAATAATTAATAATCAA 560  
 Db 599 AAAAAAATATTAATAATTAATAATTAATAATTAATAATTAATAATTAATAATTAATAAT 658  
 QY 561 GGTAAATTCGCTTTTAAATTAACATTAAGCATTAACAATTTTAAATTAACGAAGTTTATTTC 620  
 Db 659 TATTATAATTTTAAATAATTAATAATTAATAATTAATAATTAATAATTAATAATTAAT 717  
 QY 621 AGCATGGGACATATAATTAACCTTAAATTTAAATTTAAATTTGAAGATGCTTTTAAAGTTA 680  
 Db 718 -TTATTTTAAATTAATTAATAATTAATAATTAATAATTAATAATTAATAATTAATAAT 776  
 QY 681 AAGACGAGCCATACCTTATTTTCAAGCTTATTAAGCTTGACACAGGTACACTAGTCTTTT 740  
 Db 777 ATTATATATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 836  
 QY 741 ATTATTAATTTTCTGAGAAATCAAGTTTACGATCATATAATTTT 786  
 Db 837 TTTATNATAATTAATAATTAATAATTAATAATTTTAAAAAATAAATTTT 882

## RESULT 14

CD325190/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

CD325190 1123 bp mRNA linear EST 28-MAY-2003  
 AGENCOURT 14148241 NICHED\_XGC Eye1 Xenopus laevis cDNA clone  
 IMAGE:6949287 5', mRNA sequence.  
 CD325190  
 CD325190.1 GI:31089137  
 EST.  
 Xenopus laevis (African clawed frog)  
 Xenopus laevis  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;  
 Xenopodinae; Xenopus.  
 1 (bases 1 to 1123)  
 NIH-MGC <http://mgs.nci.nih.gov/>.  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished  
 Contact: Daniela S. Gerhard, Ph.D.  
 Office of Cancer Genomics  
 National Cancer Institute / NIH  
 Bldg. 31 Rm10A07 Bethesda, MD 20892  
 Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)  
 Tissue Procurement:  
 cDNA Library Preparation: Life Technologies, Inc.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
 Plate: LIA4M4564 row: 0 column: 14  
 High quality sequence start: 49  
 High quality sequence stop: 150.  
 Location/Qualifiers  
 1..1123  
 /organism="Xenopus laevis"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:8355"  
 /clone="IMAGE:6949287"  
 /dev\_stage="adult"  
 /lab\_host="DH10B (phage-resistant)"  
 /clone\_lib="NICHED\_XGC Eye1"  
 /notes="Organ: eye; Vector: pCMV-SPORT6; Site 1: NotI;  
 Site 2: Sall; Cloned unidirectionally. Primer: Oligo dt.  
 Average insert size 2.3 Kb. Constructed by Life

## FEATURES

source

RESULT 15  
 BX437758/c  
 LOCUS  
 DEFINITION  
 ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM

BX437758 Homo sapiens THYMUS  
 5-PRIME, mRNA sequence.  
 BX437758  
 BX437758.1 GI:30773605  
 EST.  
 Homo sapiens (human)  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Technologies. Note: This is a Xenopus Gene Collection (XGC  
 ) library."

BASE COUNT 248 a 65 c 49 g 662 t 99 others  
 ORIGIN

Query Match 9.7%; Score 92.2; DB 14; Length 1123;  
 Best Local Similarity 47.9%; Pred. No. 0.0041;  
 Matches 348; Conservative 0; Mismatches 368; Indels 11; Gaps 3;

QY 27 AAAAAAGTTTGGGCAATTTTTCAAAAATAAAAAATATCAATAAGTTGGAGTCATTACCG 86

Db 975 AATAAGTAAGCAAAATAGAAAAAATANAATAATAATAATAATAATAAGAAAAAATTT 916

QY 87 AATTTTATACITATTGTTTGAATGAACCTTTATAACATAGTTGGATAGAGTTTCGAT 146

Db 915 AATAACAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAA 856

QY 147 TTAATACATTAAATGTAACCTTGCTTACAACAAGATGTCATCAGAAGGAGTGGTTTAA 206

Db 855 GAAATATAAATATAT-ATATTTATCATAAAAAATAAAGTTAAAAAAGAAATAAAT 797

QY 207 AATGAGTAAAAATTAATGATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 266

Db 796 AA-----AAAATNCATTAATAATAATAATAATAATAATAATAATAATAATAA 742

QY 267 TTTCAAGATACAAAAAAGAGTTCAATTTTGAACCTATGAAGAAATTTATATTAATCA 326

Db 741 ANAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAA 682

QY 327 TATTTTAAAGTGAAGTCTTAACGAAATCTCATCTAAAGAGATTTGCTAGTCTCAGAGTT 386

Db 681 AATAAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAA 622

QY 387 CAACCTTACTATTTTAACTTAAGCTTTTACAAAAGCTTAAAGATTTTAAATTTGTTATCAA 446

Db 621 AAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAA 567

QY 447 GAAAAAGAGTTTACAGAGCAAGAACAGTTATTTTATGTTTACAGATACACAAAAAGC 506

Db 566 AATAATATAATTTTAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAA 507

QY 507 AATATTTCAAAAAGTGAATTTAGAGATTAAGAGATTAAGAGATTAAGAGATTAAGAG 566

Db 506 AAAAAAATAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAA 447

QY 567 TTGCGTTTAAATAACATTTGAAGCATTAACAAATTTTAAATACGAAGTTATTTTATTCAGCAT 626

Db 446 ATAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAA 387

QY 627 GGGACATAAAATTAACCTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAAG 686

Db 386 TATATAAAAAATATATTTAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAA 327

QY 687 AGCCATACCTTATTTTCAAGCTTATTAAGCTTGACACAGGTTACACTAGTCTTTTATTTA 746

Db 326 ATAAATAATAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAA 267

QY 747 ATATTTT 753

Db 266 TTAATTT 260





**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 17, 2003, 11:24:27 ; Search time 49 Seconds  
(without alignments)  
372.522 Million cell updates/sec

Title: US-10-043-539A-2

Perfect score: 574

Sequence: 1 MSKINDINDLVNATFQVKF.....DTQKANTQKLISEEYIKN 115

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq 19Jun03.\*

```
1: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
2: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
3: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
4: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
5: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
6: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
7: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
8: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
9: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
10: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
11: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
24: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	496	84.7	128	23	ABP40148
2	148	25.8	124	18	AAW14000
3	148	25.8	124	21	AAV50721
4	148	25.8	124	21	AAV50725
5	148	25.8	124	22	AB368760
6	146	25.4	113	21	AAV50726
7	145	25.3	149	23	ABF40306
8	143	24.9	124	22	AB68761
9	143	24.9	148	23	ABP40769

10	99.5	17.3	249	23	ABP40173	Staphylococcus epi
11	94.5	16.5	154	23	ABP40736	Staphylococcus epi
12	93.5	16.3	134	23	ABP54930	S. aureus regulato
13	93.5	16.3	144	23	ABP54930	Lactococcus lactis
14	93.5	16.3	147	23	ABP54929	Staphylococcus aur
15	92.5	16.1	141	23	ABP40135	Listeria monocytog
16	88.5	15.4	136	23	ABP40135	Staphylococcus epi
17	87	15.2	150	23	ABP40135	Listeria monocytog
18	85	14.8	558	21	ABP40135	Plasmodium falcipa
19	84.5	14.7	201	19	AAW43423	Rat syntaxin 5 pro
20	83.5	14.5	156	23	ABP38889	Staphylococcus epi
21	83.5	14.5	1003	23	AAW43045	Human ORFX ORF2809
22	83.5	14.5	1068	22	AAW78487	Human protein SEQ
23	83.5	14.5	1084	22	AAW78471	Human protein SEQ
24	81.5	14.2	1639	19	AAW54145	P. falciparum synt
25	81.5	14.2	1639	23	AAE29345	Plasmodium falcipa
26	81.5	14.2	1584	6	AAE50777	Sequence of the P1
27	81.5	14.2	2349	24	AAO16359	Human translocated
28	81.5	14.2	2400	22	ABG20278	Novel human diagno
29	81.5	14.2	2415	22	ABG20279	Novel human diagno
30	81	14.1	1176	21	AAW18280	Plasmodium falcipa
31	80.5	14.0	394	23	ABP78326	Amino acid sequenc
32	79.5	13.9	180	23	ABP39576	Staphylococcus epi
33	79.5	13.9	428	20	AAW5086	S. pneumoniae resp
34	79.5	13.9	428	20	AAW5086	S. pneumoniae resp
35	79.5	13.9	428	24	ABU00491	S. pneumoniae type
36	79	13.8	365	21	AAW41033	Human ORFX ORF797
37	79	13.8	427	21	AAW97292	Lipid associated p
38	79	13.8	567	22	AAW00976	Human bone marrow
39	78.5	13.7	134	23	ABP40764	Staphylococcus epi
40	78.5	13.7	472	23	ABT10468	XisA recombinase p
41	78.5	13.7	479	23	ABT10470	XisA-XisA fusion pr
42	78.5	13.7	1145	22	ABG2169	NLS-XisA fusion pr
43	78.5	13.7	1151	23	ABP38397	S. epidermidis ope
44	78.5	13.7	1151	23	ABP38397	Staphylococcus epi
45	78	13.6	148	23	ABP73809	Candida albicans e
						Lactococcus lactis

#### ALIGNMENTS

##### RESULT 1

ABP40148  
ID ABP40148 standard; Protein; 128 AA.

XX ABP40148;

XX AC

XX 24-JUL-2002 (first entry)

DE Staphylococcus epidermidis ORF amino acid sequence SEQ ID NO:4993.

XX Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;

XX antibacterial; gene therapy.

XX Staphylococcus epidermidis.

XX US6380370-B1.

XX 30-APR-2002.

XX 13-AUG-1998; 98US-0134001.

XX 14-AUG-1997; 97US-055779P.

XX 08-NOV-1997; 97US-064964P.

XX (GENO-) GENOME THERAPEUTICS CORP.

XX Doucette-Stamm LA, Bush D;

XX WPI: 2002-381255/41.

XX N-PSDB; AEN92693.

XX Novel isolated nucleic acid encoding a Staphylococcus epidermidis

PT polypeptide, useful for diagnosing and treating bacterial infections -  
 PS Disclosure; SEQ ID 4993; 267pp; English.  
 CC ABN90538 to ABN93374 represent Staphylococcus epidermidis open reading  
 CC frame (ORF) nucleic acid sequences which encode the amino acid sequences  
 CC given in ABP35124 to ABP37960. The S. epidermidis sequences have  
 CC antibacterial activity and can be used in gene therapy. The sequences  
 CC can also be used in the diagnosis and treatment of bacterial infections,  
 CC particularly S. epidermidis infections. The sequences can be used to  
 CC screen for compounds able to interfere with the S. epidermidis life  
 CC cycle or inhibit S. epidermidis infection.  
 CC N.B. The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from the  
 CC USPTO web site.  
 XX  
 SQ Sequence 128 AA;  
 Query Match 84.7%; Score 486; DB 23; Length 128;  
 Best Local Similarity 84.2%; Pred. No. 3.2e-40;  
 Matches 96; Conservative 8; Mismatches 10; Indels 0; Gaps 0;  
 QY 1 MSKINDINDLVNATQVKKFFRDTKKFNLYVEIYILNHILRSNESISKEIATCSEF 60  
 Db 15 MGKINDINDLVNATQVKKFFRDTKKQNLNVEIYILNHILRSNESISKEIATCSEF 74  
 QY 61 KPYLTALQKLKDLKLSKRSLOQDERTVIVYVTDTKANIQKLISELEYIK 114  
 Db 75 KPYLTALQKLKDLNLSKRSVHDERTVIVFVSDQREKIKKLIILELNYIK 128  
 RESULT 2  
 AA014000  
 ID AA014000 standard; Protein; 124 AA.  
 AC AA014000;  
 DT 23-MAY-1997 (first entry)  
 DE Staphylococcal accessory regulatory A protein.  
 XX  
 KW Staphylococcal accessory regulatory A protein; sar; virulence factor;  
 KW bacterial endotoxin; regulatory protein; exoprotein gene; inhibitor;  
 KW antimicrobial agent.  
 XX  
 CS Staphylococcus aureus.  
 XX  
 PN US5587288-A.  
 PD 24-DEC-1996.  
 PF 24-MAY-1994; 94US-0248505.  
 PR 24-MAY-1994; 94US-0248505.  
 XX (UYRQ) UNIV ROCKEFELLER.  
 XX  
 XX Cheung A, Fischetti VA;  
 XX WPI; 1997-064792/06.  
 DR N-PSDB; AAT60281.  
 XX  
 CC New isolated staphylococcal accessory regulatory protein and gene  
 CC used to develop prods. for use as antimicrobial agents and for  
 CC detection of pathogenic staphylococci.  
 XX  
 PS Claim 1; Column 9-10; 9pp; English.  
 XX  
 CC This sequence represents the staphylococcal accessory regulatory (SAR) A  
 CC protein of Staphylococcus aureus. The SAR protein controls the  
 CC expression of potential virulence factors, such as bacterial endotoxins.  
 CC The SAR locus is thought to regulate exoprotein genes positively, at the  
 CC mRNA level. This protein can be used for designing analogues which

CC interfere with the expression of the toxins, thereby functioning as  
 CC antimicrobial agents to render the microorganism avirulent. The DNA  
 CC encoding this sequence can be used for diagnostic identification of  
 CC pathogenic staphylococci which express the protein, and the resulting  
 CC virulence factor.  
 XX  
 SQ Sequence 124 AA;  
 Query Match 25.8%; Score 148; DB 18; Length 124;  
 Best Local Similarity 30.1%; Pred. No. 5.6e-07;  
 Matches 34; Conservative 27; Mismatches 52; Indels 0; Gaps 0;  
 QY 1 MSKINDINDLVNATQVKKFFRDTKKFNLYVEIYILNHILRSNESISKEIATCSEF 60  
 Db 3 ITKINDCFELLSNVTVYADKLSLIKKEFSISFEFAVLTYISENKEKEYLKDINHLY 62  
 QY 61 KPYLTALQKLKDLKLSKRSLOQDERTVIVYVTDTKANIQKLISELEYIK 113  
 Db 63 KOPVVVAVKILSQEDYDFKKENHDEHRTVLILVNAQQRKIESLLSRVNRKI 115  
 RESULT 3  
 AAY50721  
 ID AAY50721 standard; Protein; 124 AA.  
 XX  
 AC AAY50721;  
 DT 08-FEB-2000 (first entry)  
 DE S. aureus sarA protein.  
 XX  
 KW Staphylococcal accessory regulatory protein; sar protein; endotoxin;  
 KW exoprotein virulence determinant; EVD; antibacterial agent; detection.  
 XX  
 OS Staphylococcus aureus.  
 XX  
 PN US5976792-A.  
 PD 02-NOV-1999.  
 PF 08-JUL-1996; 96US-0676782.  
 PR 24-MAY-1994; 94US-0248505.  
 XX (SIGA-) SIGA PHARM INC.  
 XX  
 XX Cheung A, Fischetti VA;  
 XX WPI; 2000-021938/02.  
 DR N-PSDB; AA224204.  
 XX  
 CC New accessory regulatory protein, sar, from Staphylococcus aureus, used  
 CC to design analogs potentially useful as antibacterial agents -  
 XX  
 PS Disclosure; Column 25-28; 30pp; English.  
 XX  
 CC This invention describes a novel full-length Staphylococcus aureus  
 CC accessory regulatory protein ((II), designated sar) which regulates the  
 CC expression of S. aureus exoprotein virulence determinants (EVD). (I)  
 CC controls the expression of virulence determinants such as endotoxins  
 CC in S. aureus. (I) is used to design analogs that interfere with  
 CC expression of EVD, i.e. potential antibacterial agents and for generating  
 CC specific antibodies which are used to detect (I) in microbial isolates or  
 CC for affinity purification of (I). The nucleic acid (II) that encodes (I)  
 CC (or its fragments) can be used to identify S. aureus that express sar  
 CC (and thus EVD) by usual hybridization and amplification tests, also for  
 CC recombinant production of (I). This sequence represents the  
 CC Staphylococcus aureus sarA protein which is described in the method of  
 CC the invention.  
 XX  
 SQ Sequence 124 AA;  
 Query Match 25.8%; Score 148; DB 21; Length 124;

```

Best Local Similarity 30.1%; Pred. No. 5.6e-07;
Matches 34; Conservative 27; Mismatches 52; Indels 0; Gaps 0;

QY 1 MSKINDINDLVNATFOVKKFFRDTKKFNLYEYIILNHLRSENEISSKEIAKSEF 60
DB 3 ITKINDCFELLSMVTYADKLSLIKKEFSISFEFAVLTYISENKEEYILKDIINHLN 62
QY 61 KPYVLTAKLQKLDKLSKRSLODERTVIVVDTQKANIQKLISELEEVY 113
DB 63 KQOVVKAVKILSQEDYFDKRNEDERTVLILVNAQQRKKIESLLSRVVKRI 115

RESULT 4
AA50725
ID AAY50725 standard; Protein; 124 AA.
XX
AC AAY50725;
XX
XX 08-FEB-2000 (first entry)
XX
DE S. aureus sarA protein fragment from strain RN450.
XX
KW Staphylococcal accessory regulatory protein; sar protein; endotoxin;
XX exoprotein virulence determinant; EVD; antibacterial agent; detection.
XX
OS Staphylococcus aureus.
XX
XX US976792-A.
XX
XX 02-NOV-1999.
XX
XX 08-JUL-1996; 96US-0676782.
XX
XX 24-MAY-1994; 94US-0248505.
XX
XX (SIGA-) SIGA PHARM INC.
XX
XX Cheung A, Fischetti VA;
XX
XX WPI; 2000-021938/02.
XX
XX New accessory regulatory protein, sar, from Staphylococcus aureus, used
XX to design analogs potentially useful as antibacterial agents -
XX
XX Claim 4; Column 31-32; 30pp; English.
XX
XX This invention describes a novel full-length Staphylococcus aureus
XX accessory regulatory protein ((I), designated sar) which regulates the
XX expression of S. aureus exoprotein virulence determinants (EVD). (I)
XX controls the expression of virulence determinants such as endotoxins
XX in S. aureus. (I) is used to design analogs that interfere with
XX expression of EVD, i.e. potential antibacterial agents and for generating
XX specific antibodies which are used to detect (I) in microbial isolates or
XX for affinity purification of (I). The nucleic acid (II) that encodes (I)
XX (or its fragments) can be used to identify S. aureus that express sar
XX (and thus EVD) by usual hybridization and amplification tests, also for
XX recombinant production of (I). This sequence represents a Staphylococcus
XX aureus sarA protein which is described in the method of the invention.
XX
XX Sequence 124 AA;
XX
Query Match 25.8%; Score 148; DB 21; Length 124;
Best Local Similarity 30.1%; Pred. No. 5.6e-07;
Matches 34; Conservative 27; Mismatches 52; Indels 0; Gaps 0;

QY 1 MSKINDINDLVNATFOVKKFFRDTKKFNLYEYIILNHLRSENEISSKEIAKSEF 60
DB 3 ITKINDCFELLSMVTYADKLSLIKKEFSISFEFAVLTYISENKEEYILKDIINHLN 62
QY 61 KPYVLTAKLQKLDKLSKRSLODERTVIVVDTQKANIQKLISELEEVY 113
DB 63 KQOVVKAVKILSQEDYFDKRNEDERTVLILVNAQQRKKIESLLSRVVKRI 115

RESULT 6
AA50726
ID AAY50726 standard; Protein; 113 AA.
XX
AC AAY50726;
XX
XX 08-FEB-2000 (first entry)
XX
DE S. epidermidis sarA protein fragment.
XX
XX Staphylococcal accessory regulatory protein; sar protein; endotoxin;

```

```

RESULT 5
AAB68760
ID AAB68760 standard; Protein; 124 AA.
XX
AC AAB68760;
XX
XX 17-APR-2001 (first entry)
XX
DE Staphylococcus aureus SarA protein.
XX
KW Staphylococcus aureus; SarA; staphylococcal accessory regulator A;
XX agr; accessory gene regulator; antibacterial; SarA inhibitor;
XX virulence gene; staphylococcal infection.
XX
OS Staphylococcus aureus.
XX
XX WO200103686-A2.
XX
XX 18-JAN-2001.
XX
XX 07-JUL-2000; 2000WO-US18525.
XX
XX 08-JUL-1999; 99US-0142793.
XX
XX (UYAR-) UNIV ARKANSAS.
XX
XX Hurlburt BK, Smeitzer MS, Reichtin TW;
XX
XX WPI; 2001-112567/12.
XX
XX Identifying inhibitors of staphylococcal SarA (accessory regulator)
XX which are useful for treating staphylococcal infections, comprises
XX using specific binding sites of SarA protein on an accessory gene
XX regulator locus -
XX
XX Example; Fig 10; 79pp; English.
XX
XX The present sequence is given in a specification relating to a method for
XX identifying inhibitors of SarA (staphylococcal accessory regulator)
XX function involved in the expression of Staphylococcal virulence genes.
XX The method comprises contacting a candidate inhibitor with a SarA
XX binding site of the agr (accessory gene regulator) locus in solution
XX and assessing the binding of the candidate inhibitor to the SarA
XX binding site of the agr locus. The identified inhibitors are useful for
XX preventing and treating staphylococcal infections.
XX
XX Sequence 124 AA;
XX
Query Match 25.8%; Score 148; DB 22; Length 124;
Best Local Similarity 30.1%; Pred. No. 5.6e-07;
Matches 34; Conservative 27; Mismatches 52; Indels 0; Gaps 0;

QY 1 MSKINDINDLVNATFOVKKFFRDTKKFNLYEYIILNHLRSENEISSKEIAKSEF 60
DB 3 ITKINDCFELLSMVTYADKLSLIKKEFSISFEFAVLTYISENKEEYILKDIINHLN 62
QY 61 KPYVLTAKLQKLDKLSKRSLODERTVIVVDTQKANIQKLISELEEVY 113
DB 63 KQOVVKAVKILSQEDYFDKRNEDERTVLILVNAQQRKKIESLLSRVVKRI 115

RESULT 6
AA50726
ID AAY50726 standard; Protein; 113 AA.
XX
AC AAY50726;
XX
XX 08-FEB-2000 (first entry)
XX
DE S. epidermidis sarA protein fragment.
XX
XX Staphylococcal accessory regulatory protein; sar protein; endotoxin;

```

KW exoprotein virulence determinant; EVD; antibacterial agent; detection.  
 XX Staphylococcus epidermidis.  
 OS  
 XX US5976792-A.  
 XX  
 XX 02-NOV-1999.  
 XX  
 XX 08-JUL-1996; 96US-0676782.  
 XX  
 XX 24-MAY-1994; 94US-0248505.  
 XX  
 XX (SIGA-) SIGA PHARM INC.  
 XX  
 XX Cheung A, Fischetti VA;  
 XX WPI; 2000-021938/02.  
 XX  
 XX New accessory regulatory protein, sar, from Staphylococcus aureus, used  
 XX to design analogs potentially useful as antibacterial agents -  
 XX  
 XX Claim 4; Column 31-34; 30pp; English.  
 XX  
 XX This invention describes a novel full-length Staphylococcus aureus  
 XX accessory regulatory protein ((I), designated sar) which regulates the  
 XX expression of S. aureus exoprotein virulence determinants (EVD). (I)  
 XX controls the expression of virulence determinants such as endotoxins  
 XX in S. aureus. (I) is used to design analogs that interfere with  
 XX expression of EVD, i.e. potential antibacterial agents and for generating  
 XX specific antibodies which are used to detect (I) in microbial isolates or  
 XX for affinity purification of (I). The nucleic acid (II) that encodes (I)  
 XX (or its fragments) can be used to identify S. aureus that express sar  
 XX (and thus EVD) by usual hybridization and amplification tests, also for  
 XX recombinant production of (I). This sequence represents a Staphylococcus  
 XX aureus sarA protein which is described in the method of the invention.  
 XX  
 XX Sequence 113 AA;  
 SQ  
 Query Match 25.4%; Score 146; DB 21; Length 113;  
 Best Local Similarity 29.7%; Pred. No. 7.9e-07;  
 Matches 33; Conservative 27; Mismatches 51; Indels 0; Gaps 0;  
 QY 1 MSKINDINDLVNATFOVKKFRDTKKFNLYNEIYILNHLRSENEISSKEIAKSEF 60  
 DB 3 ITKINDCELLSWTYADKLKSLIKKFSFEFAVLTYISENKEKEYFKDIINHLNY 62  
 QY 61 KPYLTALQKLKDLKLLSKRSLODERTVIVVYDTQKANIQKLISELEE 111  
 DB 63 KQPVKAVKILSQEDYFDKXNEHDETVILVNAQQRKKIESLLSRVVK 113  
 RESULT 7  
 ABP40306  
 ID ABP40306 standard; Protein; 149 AA.  
 AC ABP40306;  
 XX  
 XX 24-JUL-2002 (first entry)  
 XX  
 XX Staphylococcus epidermidis ORF amino acid sequence SEQ ID NO:5151.  
 DE  
 XX Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;  
 KW antibacterial; gene therapy.  
 XX  
 XX Staphylococcus epidermidis.  
 OS  
 XX US6380370-B1.  
 XX  
 XX 30-APR-2002.  
 XX  
 XX 13-AUG-1998; 98US-0134001.  
 XX  
 XX 14-AUG-1997; 97US-055779P.  
 XX  
 XX Identifying inhibitors of staphylococcal SarA (accessory regulator)  
 PT which are useful for treating staphylococcal infections, comprises

PR 08-NOV-1997; 97US-064964P.  
 XX (GENO-) GENOME THERAPEUTICS CORP.  
 XX  
 XX Doucette-Stamm LA, Bush D;  
 XX WPI; 2002-381255/41.  
 DR N-PSDB; ABN92851.  
 XX  
 XX Novel isolated nucleic acid encoding a Staphylococcus epidermidis  
 PT polypeptide, useful for diagnosing and treating bacterial infections -  
 XX  
 XX Disclosure; SEQ ID 5151; 267pp; English.  
 XX  
 XX ABN90538 to ABN93374 represent Staphylococcus epidermidis open reading  
 CC frame (ORF) nucleic acid sequences which encode the amino acid sequences  
 CC given in ABP35124 to ABP37960. The S. epidermidis sequences have  
 CC antibacterial activity and can be used in gene therapy. The sequences  
 CC can also be used in the diagnosis and treatment of bacterial infections,  
 CC particularly S. epidermidis infections. The sequences can be used to  
 CC screen for compounds able to interfere with the S. epidermidis life  
 CC cycle or inhibit S. epidermidis infection.  
 CC N.B. The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from the  
 CC USPTO web site.  
 XX  
 XX Sequence 149 AA;  
 SQ  
 Query Match 25.3%; Score 145; DB 23; Length 149;  
 Best Local Similarity 33.3%; Pred. No. 1.4e-06;  
 Matches 33; Conservative 22; Mismatches 42; Indels 2; Gaps 1;  
 QY 16 QVKKFFRDTKKFNLYNEIYILNHLRSENEISSKEIAKSEF KPYLTALQKLKDL 75  
 DB 32 KIDSFNIIKKYKMSKEILILLTL--EKSWTLKEMDKYVHIKPYKRTYNNLVNL 89  
 QY 76 KLLSKRSLODERTVIVVYDTQKANIQKLISELEEYIK 114  
 DB 90 EWYKERPDDETVIIHFNDKQNSKEDLLKFIDDSIK 128  
 RESULT 8  
 AAB68761  
 ID AAB68761 standard; protein; 124 AA.  
 XX  
 XX AAB68761;  
 XX  
 XX 17-APR-2001 (first entry)  
 XX  
 XX Staphylococcus epidermidis SarA protein.  
 DE  
 XX Staphylococcus epidermidis; SarA; staphylococcal accessory regulator A;  
 KW agr; accessory gene regulator; antibacterial; SarA inhibitor;  
 KW virulence gene; staphylococcal infection.  
 XX  
 XX Staphylococcus epidermidis.  
 OS  
 XX WO200103686-A2.  
 XX  
 XX 18-JAN-2001.  
 PD  
 XX 07-JUL-2000; 2000WO-US18525.  
 XX  
 XX 08-JUL-1999; 99US-0142793.  
 XX  
 XX (UYAR-) UNIV ARKANSAS.  
 XX  
 XX Hurlburt BK, Smeltzer MS, Reichtin TM;  
 XX WPI; 2001-112567/12.  
 XX  
 XX Identifying inhibitors of staphylococcal SarA (accessory regulator)  
 PT which are useful for treating staphylococcal infections, comprises

PT using specific binding sites of SarA protein on an accessory gene  
XX regulator locus -  
PS Disclosure; Page 13; 79pp; English.  
XX  
CC The present sequence is given in a specification relating to a method for  
CC identifying inhibitors of SarA (staphylococcal accessory regulator)  
CC function involved in the expression of staphylococcal virulence genes.  
CC The method comprises contacting a candidate inhibitor with a SarA  
CC binding site of the agr (accessory gene regulator) locus in solution  
CC and assessing the binding of the candidate inhibitor to the SarA  
CC binding site of the agr locus. The identified inhibitors are useful for  
CC preventing and treating staphylococcal infections.  
XX  
SQ Sequence 124 AA;  
Query Match 24.9%; Score 143; DB 22; Length 124;  
Best Local Similarity 29.2%; Pred. No. 1.7e-06;  
Matches 33; Conservative 25; Mismatches 55; Indels 0; Gaps 0;  
QY 1 MSKINDINDLVNATFOVKKFFRDTKKKLNLYEYIILNHLRSENEISSKEIAKCSFP 60  
DB 3 ISKINDCFELLAMVTYADRLKGIKKEFSISFEFAVLTYISENKEBEYLYKDIINHLNY 62  
QY 61 KPYLYTKALQKLKDLKLSKRSLODERTVIVVYDTQKANIQKLISELEVEYI 113  
DB 63 KQPVVKA VKNL SQENYFNKRNEDERTVLILVDSKQKKIDLLKRVNRI 115  
RESULT 9  
ABP40769  
ID ABP40769 standard; Protein; 148 AA.  
XX  
AC ABP40769;  
XX  
DT 24-JUL-2002 (first entry)  
XX  
DE Staphylococcus epidermidis ORF amino acid sequence SEQ ID NO:5614.  
XX  
KW Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;  
KW antibacterial; gene therapy.  
XX  
OS Staphylococcus epidermidis.  
XX  
FN US6380370-B1.  
XX  
PD 30-APR-2002.  
XX  
PF 13-AUG-1998; 98US-0134001.  
XX  
PR 14-AUG-1997; 97US-055779P.  
PR 08-NOV-1997; 97US-064964P.  
XX  
PA (GENO-) GENOME THERAPEUTICS CORP.  
XX  
PI Doucette-Stamm LA, Bush D;  
XX  
DR WPI; 2002-381255/41.  
XX  
DR N-PSDB; ABN93314.  
XX  
PT Novel isolated nucleic acid encoding a Staphylococcus epidermidis  
PT polypeptide, useful for diagnosing and treating bacterial infections -  
XX  
PS Disclosure; SEQ ID 5614; 267pp; English.  
XX  
CC ABN90538 to ABN93374 represent Staphylococcus epidermidis open reading  
CC frame (ORF) nucleic acid sequences which encode the amino acid sequences  
CC given in ABP35124 to ABP37960. The S. epidermidis sequences have  
CC antibacterial activity and can be used in gene therapy. The sequences  
CC can also be used in the diagnosis and treatment of bacterial infections,  
CC particularly S. epidermidis infections. The sequences can be used to  
CC screen for compounds able to interfere with the S. epidermidis life  
CC cycle or inhibit S. epidermidis infection.  
XX  
PS Disclosure; SEQ ID 5614; 267pp; English.  
XX  
CC ABN90538 to ABN93374 represent Staphylococcus epidermidis open reading  
CC frame (ORF) nucleic acid sequences which encode the amino acid sequences  
CC given in ABP35124 to ABP37960. The S. epidermidis sequences have  
CC antibacterial activity and can be used in gene therapy. The sequences  
CC can also be used in the diagnosis and treatment of bacterial infections,  
CC particularly S. epidermidis infections. The sequences can be used to  
CC screen for compounds able to interfere with the S. epidermidis life  
CC cycle or inhibit S. epidermidis infection.

CC N.B. The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from the  
CC USPTO web site.  
XX  
SQ Sequence 148 AA;  
Query Match 24.9%; Score 143; DB 23; Length 148;  
Best Local Similarity 29.2%; Pred. No. 2.2e-06;  
Matches 33; Conservative 25; Mismatches 55; Indels 0; Gaps 0;  
QY 1 MSKINDINDLVNATFOVKKFFRDTKKKLNLYEYIILNHLRSENEISSKEIAKCSFP 60  
DB 27 ISKINDCFELLAMVTYADRLKGIKKEFSISFEFAVLTYISENKEBEYLYKDIINHLNY 86  
QY 61 KPYLYTKALQKLKDLKLSKRSLODERTVIVVYDTQKANIQKLISELEVEYI 113  
DB 87 KQPVVKA VKNL SQENYFNKRNEDERTVLILVDSKQKKIDLLKRVNRI 139  
RESULT 10  
ABP40173  
ID ABP40173 standard; Protein; 249 AA.  
XX  
AC ABP40173;  
XX  
DT 24-JUL-2002 (first entry)  
XX  
DE Staphylococcus epidermidis ORF amino acid sequence SEQ ID NO:5018.  
XX  
KW Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;  
KW antibacterial; gene therapy.  
XX  
OS Staphylococcus epidermidis.  
XX  
FN US6380370-B1.  
XX  
PD 30-APR-2002.  
XX  
PF 13-AUG-1998; 98US-0134001.  
XX  
PR 14-AUG-1997; 97US-055779P.  
PR 08-NOV-1997; 97US-064964P.  
XX  
PA (GENO-) GENOME THERAPEUTICS CORP.  
XX  
PI Doucette-Stamm LA, Bush D;  
XX  
DR WPI; 2002-381255/41.  
XX  
DR N-PSDB; ABN92718.  
XX  
PT Novel isolated nucleic acid encoding a Staphylococcus epidermidis  
PT polypeptide, useful for diagnosing and treating bacterial infections -  
XX  
PS Disclosure; SEQ ID 5018; 267pp; English.  
XX  
CC ABN90538 to ABN93374 represent Staphylococcus epidermidis open reading  
CC frame (ORF) nucleic acid sequences which encode the amino acid sequences  
CC given in ABP35124 to ABP37960. The S. epidermidis sequences have  
CC antibacterial activity and can be used in gene therapy. The sequences  
CC can also be used in the diagnosis and treatment of bacterial infections,  
CC particularly S. epidermidis infections. The sequences can be used to  
CC screen for compounds able to interfere with the S. epidermidis life  
CC cycle or inhibit S. epidermidis infection.  
XX  
PS Disclosure; SEQ ID 5018; 267pp; English.  
XX  
CC N.B. The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from the  
CC USPTO web site.  
XX  
SQ Sequence 249 AA;  
Query Match 17.3%; Score 99.5; DB 23; Length 249;  
Best Local Similarity 28.8%; Pred. No. 0.078;  
Matches 23; Conservative 20; Mismatches 36; Indels 1; Gaps 1;

and  
terial,  
terial

33  
33

aps LTKA 68  
::  
VSPL 72



DT 08-JAN-2003 (first entry)  
XX Staphylococcus aureus regulator of autolytic activity (RAT).  
XX Regulator of autolytic activity; RAT, autolysis; antimicrobial;  
KW antibacterial; screening.  
XX Staphylococcus aureus.  
OS  
XX WO200270666-A2.  
XX 12-SEP-2002.  
XX 06-MAR-2002; 2002WO-US06844.  
XX 06-MAR-2001; 2001US-273791P.  
PR 15-AUG-2001; 2001US-312546P.  
PR 12-OCT-2001; 2001US-329140P.  
XX (DART-) DARTMOUTH COLLEGE.  
XX  
XX Cheung A;  
XX WPI; 2002-750455/81.  
DR N-PSDB; ABV73907.  
XX  
XX New regulator of autolytic activity (RAT) nucleic acid sequences and  
PT polypeptides, useful for regulating the autolytic activity of bacteria,  
PT inhibiting bacterial growth and infectivity, or for treating bacterial  
PT infection -  
XX  
XX Claim 7; Page 25; 26pp; English.  
XX The present sequence is a novel regulatory of autolytic activity  
CC (RAT) protein of Staphylococcus aureus, which is involved in  
CC the control, modulation or regulation of autolytic activity. This  
CC 17 kDa protein (predicted pI 7.38) is encoded by a novel gene (see  
CC ABV73907) that was identified in a Tn551 transposon library of S.  
CC aureus strain RN6390 by screening for genes that affected expression  
CC of genes encoding the capsular polysaccharide of S. aureus. The  
CC invention provides RAT gene and RAT mutant nucleic acid sequences,  
CC the polypeptides encoded by these genes, vectors and host cells,  
CC and a method of identifying agents that modulate the autolytic,  
CC activity of S. aureus and other bacteria through interaction with  
CC the RAT gene, RAT mutant gene, or the polypeptides encoded by them.  
CC These agents are expected to be useful in the inhibition of growth  
CC of S. aureus and other bacteria and in the treatment of bacterial  
CC infection. The RAT gene or gene products are also useful as  
CC research tools to enhance the understanding of the autolytic  
CC process of bacteria.  
XX  
XX Sequence 147 AA;  
XX SQ  
Query Match 16.3%; Score 93.5; DB 23; Length 147;  
Best Local Similarity 28.4%; Pred. No. 0.16;  
Matches 25; Conservative 25; Mismatches 35; Indels 3; Gaps 3  
QY 10 LVNATFQVKFFRD-TKKKFNLYEEIYILNHLRSESSNEISSKEIKACSEFKYYITKA 68  
DB 15 LYNQQRVRYSNKVFKNLYTPQFLVLT-ILWDES-PVNVKVVTEALDGTGVSPL 72  
QY 69 LQKLDKLLSKRSLODERTVIVVVD 96  
DB 73 LARMEQVDLKRERSEVDQREVFHILD 100  
RESULT 15  
ABB48360  
ID ABB48360 standard; Protein; 141 AA.  
XX  
AC ABB48360;  
XX  
DT 05-FEB-2002 (first entry)  
73 LKRMQVDLKRERSEVDQREVFHILD 100  
RESULT 13  
ABB55470  
ID ABB55470 standard; Protein; 144 AA.  
XX  
AC ABB55470;  
XX  
DT 16-MAY-2002 (first entry)  
XX Lactococcus lactis protein zitR.  
DE Lactococcus lactis protein zitR.  
XX  
KW Biosynthesis; biodegradation; lactic bacterium; yogurt; cheese.  
XX  
OS Lactococcus lactis IL1403.  
XX  
XX FR2807446-A1.  
XX  
PD 12-OCT-2001.  
XX  
XX 11-APR-2000; 2000FR-0004630.  
XX  
XX 11-APR-2000; 2000FR-0004630.  
XX  
XX (INRG ) INRA INST NAT RECH AGRONOMIQUE.  
XX  
XX Bolotine A, Sorokine A, Renault P, Ehrlich SD;  
XX  
XX WPI; 2002-043418/06.  
DR  
XX New nucleotide sequence useful in the identification of Lactococcus  
PT lactis and related species -  
PT  
XX  
XX Claim 6; SEQ ID No 2172; 2504pp; French.  
XX  
XX The present invention is related to a Lactococcus lactis nucleotide  
CC sequence (ABA90521) and related proteins (ABB53300-ABB55621). The  
CC nucleic acid sequence is useful in the detection and/or amplification of  
CC nucleic acid sequence, particularly to identify Lactococcus lactis or  
CC related species. The proteins of the invention are useful for the  
CC biosynthesis or biodegradation of a composition of interest. The  
CC invention helps research in lactic bacteria, particularly useful in the  
CC production of yogurt and cheese.  
CC Note: The sequence data for this patent is based on equivalent patent  
CC WO2001/77334 (published 18-OCT-2001) which is available in electronic  
CC format directly from WIPO at ftp.wipo.int/pub/published\_pat\_sequences.  
XX  
XX Sequence 144 AA;  
XX SQ  
Query Match 16.3%; Score 93.5; DB 23; Length 144;  
Best Local Similarity 27.4%; Pred. No. 0.16;  
Matches 34; Conservative 23; Mismatches 50; Indels 17; Gaps 4;  
QY 1 MSKINDINDLVNATFQVKFFRDTKKFNLYEEIYI-----LNHLRSESSNEISSK-EI 54  
DB 1 NSLANQIDQFLGTIMQ-----FAENKHELLJGKCSVDVKLTSTQEHILLMLAEQISTNAKI 56  
QY 55 AKCSGFEKPYITKALQKLDKLLSKRSLODERTVI-----VYVTDQKANIQKLIS 107  
DB 57 AEKLKISPAATVKALKLQEQELIKSSRATNDRVVLWSLTERKAVPVAKETHHEKTLIS 116  
QY 108 ELEE 111  
DB 117 TYQE 120  
RESULT 14  
ABP54929  
ID ABP54929 standard; Protein; 147 AA.  
XX  
AC ABP54929;  
XX

Search completed: November 17, 2003, 12:43:13  
Job time : 51 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 17, 2003, 12:42:30 , Search time 21 Seconds  
(without alignments)  
231.702 Million cell updates/sec

Title: US-10-043-539A-2  
Perfect score: 574  
Sequence: 1 MSKINDINDLVNATFQVKF.....DTQKANIQLISEEYIKN 115

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents, AA.\*  
1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pep.\*  
2: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep.\*  
3: /cgn2\_6/ptodata/2/iaa/6A\_COMB.pep.\*  
4: /cgn2\_6/ptodata/2/iaa/6B\_COMB.pep.\*  
5: /cgn2\_6/ptodata/2/iaa/PTUS\_COMB.pep.\*  
6: /cgn2\_6/ptodata/2/iaa/backfiles.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	486	84.7	128	4	US-09-134-001C-4993
2	148	25.8	124	4	US-08-676-782-12
3	146	25.4	113	2	US-08-676-782-13
4	145	25.3	149	4	US-09-134-001C-5151
5	143	24.9	148	4	US-09-134-001C-5614
6	99.5	17.3	249	4	US-09-134-001C-5018
7	94.5	16.5	154	4	US-09-134-001C-5581
8	90.5	15.8	150	4	US-09-328-352-8073
9	88.5	15.4	136	4	US-09-134-001C-4980
10	84.5	14.7	301	1	US-08-393-985-12
11	84	14.6	368	4	US-09-252-991A-26170
12	83.5	14.5	156	4	US-09-134-001C-3734
13	83	14.5	152	4	US-09-107-532A-6399
14	80.5	14.0	394	4	US-08-195-705-4
15	79.5	13.9	180	4	US-09-134-001C-4421
16	79.5	13.9	428	3	US-08-138-064-2
17	79.5	13.9	428	3	US-08-138-064-4
18	79.5	13.9	428	3	US-09-487-370-2
19	78.5	13.9	428	3	US-09-487-370-4
20	78.5	13.7	134	4	US-09-134-001C-5609
21	78.5	13.7	1151	4	US-09-134-001C-3242
22	78	13.6	2285	4	US-09-308-375-2
23	77.5	13.5	516	4	US-08-107-532A-7156
24	76.5	13.3	394	4	US-08-195-705-2
25	76.5	13.3	2662	4	US-09-595-684B-31
26	76	13.2	1388	4	US-09-572-191-2
27	76	13.2	1388	4	US-09-723-262-2

28	76	13.2	1388	4	US-09-723-219-2	Sequence 2, Appli
29	74.5	13.0	275	4	US-09-107-532A-6512	Sequence 6512, Ap
30	74	12.9	574	4	US-09-134-001C-5409	Sequence 5409, Ap
31	73.5	12.8	682	4	US-09-107-532A-6403	Sequence 6403, Ap
32	73.5	12.8	1786	3	US-08-973-462-8	Sequence 8, Appli
33	73	12.7	405	4	US-09-134-001C-5259	Sequence 5259, Ap
34	73	12.7	1296	1	US-08-480-604A-28	Sequence 28, Appli
35	73	12.7	1296	2	US-08-405-496A-28	Sequence 28, Appli
36	73	12.7	1296	3	US-08-915-136-28	Sequence 28, Appli
37	72	12.5	893	1	US-07-977-434-4	Sequence 4, Appli
38	72	12.5	893	1	US-08-458-819-4	Sequence 4, Appli
39	72	12.5	893	3	US-09-105-697-10	Sequence 10, Appli
40	72	12.5	893	5	PCN-US91-07035-4	Sequence 4, Appli
41	71.5	12.5	534	4	US-09-508-370A-5	Sequence 5, Appli
42	71.5	12.5	804	4	US-09-107-532A-6348	Sequence 6348, Ap
43	71	12.4	154	4	US-09-134-001C-4123	Sequence 4123, Ap
44	71	12.4	197	4	US-09-328-352-7414	Sequence 7414, Ap
45	71	12.4	338	3	US-08-722-184-8	Sequence 8, Appli

ALIGNMENTS

RESULT 1  
US-09-134-001C-4993  
; Sequence 4993, Application US/09134001C  
; Patent No. 6380370  
; GENERAL INFORMATION:  
; APPLICANT: Lynn Doucette-Stamm et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
; FILE REFERENCE: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: GTC-007  
; CURRENT APPLICATION NUMBER: US/09/134,001C  
; CURRENT FILING DATE: 1998-08-13  
; PRIOR APPLICATION NUMBER: US 60/064,964  
; PRIOR FILING DATE: 1997-11-08  
; PRIOR APPLICATION NUMBER: US 60/055,779  
; PRIOR FILING DATE: 1997-08-14  
; NUMBER OF SEQ ID NOS: 5674  
; SEQ ID NO 4993  
; LENGTH: 128  
; TYPE: PRT  
; ORGANISM: Staphylococcus epidermidis  
US-09-134-001C-4993

Query Match	84.7%	Score	486	DB	4	Length	128
Best Local Similarity	84.2%	Pred. No.	5.9e-44				
Matches	96	Conservative	8	Mismatches	10	Indels	0
						Gaps	0
Qy	1	MSKINDINDLVNATFQVKFFRDTKKFNLYVEEYILNHLRSNEISSKEIACSEF	60				
Db	15	MGRKIDINDLVNATFQVKFFRDTKKQNLNVEEYILNHLKSENEISSKEIATCSSEF	74				
Qy	61	KPYLTALQKLKDLKLLSKRSLODERTVIVVVDTKANIKLISEEYIK 114					
Db	75	KPYLTALQKLKDLNLLSKRSVHDERIVIVFVDSQREKIKLLELENYIK 128					

RESULT 2  
US-08-676-782-12  
; Sequence 12, Application US/08676782  
; Patent No. 5976792  
; GENERAL INFORMATION:  
; APPLICANT: CHEUNG, Ambrose  
; APPLICANT: FISCHETTI, Vincent A.  
; TITLE OF INVENTION: REGULATION OF EXOPROTEIN IN  
; TITLE OF INVENTION: STAPHYLOCOCCUS AUREUS  
; NUMBER OF SEQUENCES: 18  
; CORRESPONDENCE ADDRESS:  
; ADDRESS: BURNS, DOANE, SWECKER & MATHIS, L.L.P.  
; STREET: P.O. Box 1404  
; City: Alexandria  
; STATE: Virginia

/ COUNTRY: United States  
/ ZIP: 22313-1404  
/ COMPUTER READABLE FORM:  
/ MEDIUM TYPE: Floppy disk  
/ COMPUTER: IBM PC compatible  
/ OPERATING SYSTEM: PC-DOS/MS-DOS  
/ SOFTWARE: Patent in Release #1.0, Version #1.30  
/ CURRENT APPLICATION DATA:  
/ APPLICATION NUMBER: US/08/676,782  
/ FILING DATE: 08-JUL-1996  
/ CLASSIFICATION: 435  
/ PRIOR APPLICATION DATA:  
/ APPLICATION NUMBER: US 08/248,505  
/ FILING DATE: 25-MAY-1994  
/ ATTORNEY/AGENT INFORMATION:  
/ NAME: McGowan, Malcolm K.  
/ REGISTRATION NUMBER: 39,300  
/ REFERENCE/DOCKET NUMBER: 016921-092  
/ TELECOMMUNICATION INFORMATION:  
/ TELEPHONE: (703) 836-6620  
/ TELEFAX: (703) 836-2021  
/ INFORMATION FOR SEQ ID NO: 12:  
/ SEQUENCE CHARACTERISTICS:  
/ LENGTH: 124 amino acids  
/ TYPE: amino acid  
/ STRANDEDNESS: single  
/ TOPOLOGY: linear  
/ MOLECULE TYPE: protein  
/ US-08-676-782-12

Query Match 25.8%; Score 148; DB 2; Length 124;  
Best Local Similarity 30.1%; Pred. No. 2,1e-08;  
Matches 34; Conservative 27; Mismatches 52; Indels 0; Gaps 0;

QY 1 MSKINDINDLVNATQVKKFRDTRKKFNLYEYIILNHLRSENEISKEIAKCSF 60  
Db 3 ITKINDCFELLSMVTYADKLKSLIKKEFSISFEFAVLTYISENKEKEYLKDINHLY 62

QY 61 KPYLTALQKLKDLKLSKRSLODERTVIVVYDTQKANIQKLISELEYI 113  
Db 63 KQPQVVKAVKILSQEDYDFKKRNEHDERTVLILVNAQQRKKIESLLSRVVK 115

RESULT 3  
US-08-676-782-13  
Sequence 13, Application US/08676782  
Patent No. 5976792  
GENERAL INFORMATION:  
APPLICANT: CHEUNG, Ambrose  
APPLICANT: FISCHETTI, Vincent A.  
TITLE OF INVENTION: REGULATION OF EXOPROTEIN IN  
TITLE OF INVENTION: STAPHYLOCOCCUS AUREUS  
NUMBER OF SEQUENCES: 18  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS, L.L.P.  
STREET: P.O. Box 1404  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/676,782  
FILING DATE: 08-JUL-1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/248,505  
FILING DATE: 25-MAY-1994  
ATTORNEY/AGENT INFORMATION:

/ NAME: McGowan, Malcolm K.  
/ REGISTRATION NUMBER: 39,300  
/ REFERENCE/DOCKET NUMBER: 016921-092  
/ TELECOMMUNICATION INFORMATION:  
/ TELEPHONE: (703) 836-6620  
/ TELEFAX: (703) 836-2021  
/ INFORMATION FOR SEQ ID NO: 13:  
/ SEQUENCE CHARACTERISTICS:  
/ LENGTH: 113 amino acids  
/ TYPE: amino acid  
/ STRANDEDNESS: single  
/ TOPOLOGY: linear  
/ MOLECULE TYPE: protein  
/ US-08-676-782-13

Query Match 25.4%; Score 146; DB 2; Length 113;  
Best Local Similarity 29.7%; Pred. No. 3e-08;  
Matches 33; Conservative 27; Mismatches 51; Indels 0; Gaps 0;

QY 1 MSKINDINDLVNATQVKKFRDTRKKFNLYEYIILNHLRSENEISKEIAKCSF 60  
Db 3 ITKINDCFELLSMVTYADKLKSLIKKEFSISFEFAVLTYISENKEKEYLKDINHLY 62

QY 61 KPYLTALQKLKDLKLSKRSLODERTVIVVYDTQKANIQKLISELEYI 111  
Db 63 KQPQVVKAVKILSQEDYDFKKRNEHDERTVLILVNAQQRKKIESLLSRVVK 113

RESULT 4  
US-09-134-001C-5151  
Sequence 5151, Application US/09134001C  
Patent No. 6380370  
GENERAL INFORMATION:  
APPLICANT: Lynn Doucette-Stamm et al  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS  
FILE REFERENCE: GTC-007  
CURRENT APPLICATION NUMBER: US/09/134,001C  
CURRENT FILING DATE: 1998-08-13  
PRIOR APPLICATION NUMBER: US 60/064,964  
PRIOR FILING DATE: 1997-11-08  
PRIOR APPLICATION NUMBER: US 60/055,779  
PRIOR FILING DATE: 1997-08-14  
NUMBER OF SEQ ID NOS: 5674  
SEQ ID NO 5151  
LENGTH: 149  
TYPE: PRT  
ORGANISM: Staphylococcus epidermidis  
US-09-134-001C-5151

Query Match 25.3%; Score 145; DB 4; Length 149;  
Best Local Similarity 33.3%; Pred. No. 5.5e-08;  
Matches 33; Conservative 22; Mismatches 42; Indels 2; Gaps 1;

QY 16 QVKKFRDTRKKFNLYEYIILNHLRSENEISKEIAKCSFEPYVLTALQKLKDL 75  
Db 32 KIDSIFNTTKKYGSKBEILLLTL--EKGSMTLKEMDKYVHKPYRTYNNLVNL 89

QY 76 KLLSKRSLODERTVIVVYDTQKANIQKLISELEYI 114  
Db 90 EWIYKERPODDERTVIIHFNDKQNSKEDLLKFIIDDSIK 128

RESULT 5  
US-09-134-001C-5614  
Sequence 5614, Application US/09134001C  
Patent No. 6380370  
GENERAL INFORMATION:  
APPLICANT: Lynn Doucette-Stamm et al  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS  
FILE REFERENCE: GTC-007  
CURRENT APPLICATION NUMBER: US/09/134,001C





NUMBER OF SEQUENCES: 7310  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: GENOME THERAPEUTICS CORPORATION  
STREET: 100 Beaver Street  
CITY: Waltham  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02354  
COMPUTER READABLE FORM:  
MEDIUM TYPE: CD-ROM ISO9660  
COMPUTER: PC  
OPERATING SYSTEM: <Unknown>  
SOFTWARE: ASCII  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/107,532A  
FILING DATE: 30-Jun-1998  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: 60/085,598  
FILING DATE: 14 May 1998  
APPLICATION NUMBER: 60/051571  
FILING DATE: July 2, 1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Arinello, Pamela Deneka  
REGISTRATION NUMBER: 40,489  
REFERENCE/DOCKET NUMBER: GTC-012  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (781)893-8277  
TELEFAX: (781)893-8277  
INFORMATION FOR SEQ ID NO: 6399:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 152 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: YES  
ORIGINAL SOURCE:  
ORGANISM: Enterococcus faecium  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (B) LOCATION 1...152  
SEQUENCE DESCRIPTION: SEQ ID NO: 6399:  
US-09-107-532A-6399  
Query Match 14.5%; Score 83; DB 4; Length 152;  
Best Local Similarity 26.4%; Pred. No. 0.19;  
Matches 32; Conservative 25; Mismatches 52; Indels 12; Gaps 4;  
QY 3 KINDINDLVN-----ATFQVKKFFRDTKKFNLYEEIYLNLHILSESNEISSKEI 54  
Db 6 KMEDLY-LVNQVCFPLVYAAKKEIIRHYTSLKPLNLTQVLVM--LVLWEKKTSVKEL 62  
QY 55 AKCSFKPYLTALQKLKDLKLSKRSLODERTVIIVYVDTQKANIQKLISLEBYIK 114  
Db 63 GNDLCLDSGTLTPLLKLAENKYNINRQETGDERIVNVTLT-PQGAALQKQAEIGPKIX 121  
QY 115 N 115  
Db 122 S 122  
RESULT 14  
US-08-195-705-4  
Sequence 4, Application US/08195705  
Patent No. 6420523  
GENERAL INFORMATION:  
APPLICANT: Chang, Sandra  
APPLICANT: Hui, George  
APPLICANT: Barr, Philip  
APPLICANT: Gibson, Helen  
TITLE OF INVENTION: BACULOVIRUS PRODUCED PLASMIDIUM  
TITLE OF INVENTION: FALCIPARUM VACCINE  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:

ADDRESSEE: Davis Hoxie Faithfull Hapgood  
STREET: 45 Rockefeller Pl.  
CITY: New York  
STATE: N.Y.  
COUNTRY: USA  
ZIP: 10111  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/195,705  
FILING DATE: 14-FEB-1994  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Jacobs, Seth H  
REGISTRATION NUMBER: 32140  
REFERENCE/DOCKET NUMBER: 11880A3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-757-2200  
TELEFAX: 212-586-1461  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 394 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
HYPOTHETICAL: NO  
FRAGMENT TYPE: C-terminal  
ORIGINAL SOURCE:  
ORGANISM: Plasmodium falciparum  
STRAIN: MAD  
US-08-195-705-4  
Query Match 14.0%; Score 80.5; DB 4; Length 394;  
Best Local Similarity 23.8%; Pred. No. 1.2;  
Matches 36; Conservative 28; Mismatches 44; Indels 43; Gaps 8;  
QY 6 DINDLVNATQVKKFFRDTKKFNLYEEI-----VILN---HILRSSENEISSK----- 52  
Db 46 NLNDILNSRLKRRKRYFLDVLESOLMFKHISSEYIEDSFKLNSEQKNILKSVYIK 105  
QY 53 -----BIAKSEFKPYLTALQKLKDLKLSKRSLODERTVI-----VYVDT 97  
Db 106 ESVENDIKFAQEGISYV-EKVLAKYKD-DLESIKVKEEKFPSSPPTTPSPAKTDE 163  
QY 98 QK-----ANIQ-----KLISELEYIKN 115  
Db 164 QKESKFLPELTNIETLYNNLVNKNIDYLN 194  
RESULT 15  
US-09-134-001C-4421  
Sequence 4421, Application US/09134001C  
Patent No. 6380370  
GENERAL INFORMATION:  
APPLICANT: Lynn Doucette-Stamm et al  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS  
FILE REFERENCE: GTC-007  
CURRENT APPLICATION NUMBER: US/09/134,001C  
CURRENT FILING DATE: 1998-08-13  
PRIOR APPLICATION NUMBER: US 60/064,964  
PRIOR FILING DATE: 1997-11-08  
PRIOR APPLICATION NUMBER: US 60/055,779  
PRIOR FILING DATE: 1997-08-14  
NUMBER OF SEQ ID NOS: 5674  
SEQ ID NO 4421  
LENGTH: 180  
TYPE: PRT  
ORGANISM: Staphylococcus epidermidis





GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 17, 2003, 12:44:45 ; Search time 267 Seconds  
(without alignments)  
78.630 Million cell updates/sec

Title: US-10-043-539A-2  
Perfect score: 574  
Sequence: 1 MSKINDINDLVNATFQVKFF.....DTQKANTQKLISELEEYIKN 115

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 666188 seqs, 182559486 residues

Total number of hits satisfying chosen parameters: 666188

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

- Database :
- 1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*
  - 2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*
  - 3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*
  - 4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*
  - 5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*
  - 6: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep.\*
  - 7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*
  - 8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*
  - 9: /cgn2\_6/ptodata/2/pubpaa/US09A\_PUBCOMB.pep.\*
  - 10: /cgn2\_6/ptodata/2/pubpaa/US09B\_PUBCOMB.pep.\*
  - 11: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pep.\*
  - 12: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep.\*
  - 13: /cgn2\_6/ptodata/2/pubpaa/US10A\_PUBCOMB.pep.\*
  - 14: /cgn2\_6/ptodata/2/pubpaa/US10B\_PUBCOMB.pep.\*
  - 15: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep.\*
  - 16: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep.\*
  - 17: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*
  - 18: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	ID	Description
1	574	100.0	115	US-10-043-539-2
2	224	39.0	120	US-10-043-539-26
3	148	25.8	124	US-10-290-142-9
4	148	25.8	124	US-10-043-539-3
5	148	25.8	124	US-10-290-143-8
6	123	21.4	247	US-10-145-602-4
7	120	20.9	166	US-10-145-802-2
8	115	20.0	118	US-10-043-539-28
9	95	16.6	119	US-10-043-539-27
10	93.5	16.3	134	US-10-092-264-4
11	93.5	16.3	134	US-10-290-142-4
12	93.5	16.3	147	US-10-092-264-2
13	93.5	16.3	147	US-10-290-142-2
14	93.5	16.3	147	US-10-290-143-3
15	91	15.9	153	US-10-290-142-8

16	91	15.9	153	15	US-10-290-143-7	Sequence 7, Appli
17	85	14.8	150	15	US-10-290-142-5	Sequence 5, Appli
18	85	14.8	150	15	US-10-290-143-4	Sequence 4, Appli
19	83	14.5	153	15	US-10-156-761-11146	Sequence 11146, A
20	81.5	14.2	1639	15	US-10-087-464-10	Sequence 10, Appli
21	80.5	14.0	143	15	US-10-290-142-6	Sequence 6, Appli
22	80.5	14.0	143	15	US-10-290-143-5	Sequence 5, Appli
23	80.5	14.0	394	15	US-10-062-809-3	Sequence 3, Appli
24	78.5	13.7	1881	12	US-10-032-585-7646	Sequence 7646, Ap
25	78	13.6	344	11	US-09-820-843A-35	Sequence 35, Appli
26	78	13.6	1948	12	US-10-032-585-7611	Sequence 7611, Ap
27	78	13.6	2285	10	US-09-932-183A-2	Sequence 2, Appli
28	77.5	13.5	157	10	US-09-738-626-6808	Sequence 6808, Ap
29	76.5	13.3	383	12	US-10-057-531A-5	Sequence 5, Appli
30	76.5	13.3	383	12	US-10-057-532A-5	Sequence 5, Appli
31	76.5	13.3	383	14	US-10-098-514-2	Sequence 2, Appli
32	76.5	13.3	383	14	US-10-098-514-4	Sequence 4, Appli
33	76.5	13.3	391	12	US-10-057-531A-7	Sequence 7, Appli
34	76.5	13.3	391	12	US-10-057-532A-7	Sequence 7, Appli
35	76.5	13.3	393	12	US-10-057-531A-3	Sequence 3, Appli
36	76.5	13.3	393	12	US-10-057-532A-3	Sequence 3, Appli
37	76.5	13.3	394	10	US-09-978-756-3	Sequence 3, Appli
38	76.5	13.3	394	15	US-10-062-809-2	Sequence 2, Appli
39	76.5	13.3	396	15	US-10-062-809-18	Sequence 18, Appli
40	76.5	13.3	402	14	US-10-098-514-14	Sequence 14, Appli
41	76.5	13.3	402	15	US-10-062-809-16	Sequence 16, Appli
42	76.5	13.3	431	12	US-10-057-531A-2	Sequence 2, Appli
43	76.5	13.3	431	12	US-10-057-532A-2	Sequence 2, Appli
44	76.5	13.3	546	12	US-10-057-531A-1	Sequence 1, Appli
45	76.5	13.3	546	12	US-10-057-532A-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1

US-10-043-539-2  
; Sequence 2, Application US/10043539  
; Publication No. US20030114650A1  
; GENERAL INFORMATION:  
; APPLICANT: Cheung, Ambrose  
; APPLICANT: Zhang, Gongyi  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR AFFECTING VIRULENCE DETERMINANTS IN  
; TITLE OF INVENTION: BACTERIA  
; FILE REFERENCE: DC-0199  
; CURRENT APPLICATION NUMBER: US/10/043,539  
; CURRENT FILING DATE: 2002-01-11  
; PRIOR APPLICATION NUMBER: US 60/261,233  
; PRIOR FILING DATE: 2001-01-12  
; PRIOR APPLICATION NUMBER: US 60/261,607  
; PRIOR FILING DATE: 2001-01-12  
; PRIOR APPLICATION NUMBER: US 60/289,601  
; PRIOR FILING DATE: 2001-05-08  
; NUMBER OF SEQ ID NOS: 34  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 2  
; LENGTH: 115  
; TYPE: PRT  
; ORGANISM: Staphylococcus aureus  
US-10-043-539-2

Query Match	100.0%	Score 574;	DB 15;	Length 115;
Best Local Similarity	100.0%;	Pred. No. 4.6e-50;		
Matches 115;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	MSKINDINDLVNATFQVKFFRDTKKKFNLYVEEYILNHLRSENESSKEIAKCSF	60	
Db	1	MSKINDINDLVNATFQVKFFRDTKKKFNLYVEEYILNHLRSENESSKEIAKCSF	60	
Qy	61	KPYLTKALQKLDKLLSKRSLODERTVIVVYDTQKANTQKLISELEEYIKN	115	
Db	61	KPYLTKALQKLDKLLSKRSLODERTVIVVYDTQKANTQKLISELEEYIKN	115	

RESULT 2  
US-10-043-539-26  
Sequence 26, Application US/10043539  
Publication No. US20030114650A1  
GENERAL INFORMATION:  
APPLICANT: Cheung, Ambrose  
APPLICANT: Manna, Adar  
APPLICANT: Zhang, Gongyi  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR AFFECTING VIRULENCE DETERMINANTS IN  
FILE REFERENCE: DC-0199  
CURRENT APPLICATION NUMBER: US/10/043,539  
CURRENT FILING DATE: 2002-01-11  
PRIOR APPLICATION NUMBER: US 60/261,233  
PRIOR FILING DATE: 2001-01-12  
PRIOR APPLICATION NUMBER: US 60/261,607  
PRIOR FILING DATE: 2001-01-12  
PRIOR APPLICATION NUMBER: US 60/289,601  
PRIOR FILING DATE: 2001-05-08  
NUMBER OF SEQ ID NOS: 34  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 26  
LENGTH: 120  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Consensus sequence.  
FEATURE:  
NAME/KEY: MISC\_FEATURE  
LOCATION: (2)..(2)  
OTHER INFORMATION: "X" is defined as any amino acid residue.  
FEATURE:  
NAME/KEY: MISC\_FEATURE  
LOCATION: (4)..(4)  
OTHER INFORMATION: "X" is defined as any amino acid residue.  
FEATURE:  
NAME/KEY: MISC\_FEATURE  
LOCATION: (10)..(10)  
OTHER INFORMATION: "X" is defined as any amino acid residue.  
FEATURE:  
NAME/KEY: MISC\_FEATURE  
LOCATION: (15)..(17)  
OTHER INFORMATION: "X" is defined as any amino acid residue.  
FEATURE:  
NAME/KEY: MISC\_FEATURE  
LOCATION: (19)..(20)  
OTHER INFORMATION: "X" is defined as any amino acid residue.  
FEATURE:  
NAME/KEY: MISC\_FEATURE  
LOCATION: (24)..(25)  
OTHER INFORMATION: "X" is defined as any amino acid residue.  
FEATURE:  
NAME/KEY: MISC\_FEATURE  
LOCATION: (31)..(31)  
OTHER INFORMATION: "X" is defined as any amino acid residue.  
FEATURE:  
NAME/KEY: MISC\_FEATURE  
LOCATION: (38)..(38)  
OTHER INFORMATION: "X" is defined as any amino acid residue.  
FEATURE:  
NAME/KEY: MISC\_FEATURE  
LOCATION: (44)..(45)  
OTHER INFORMATION: "X" is defined as any amino acid residue.  
FEATURE:  
NAME/KEY: MISC\_FEATURE  
LOCATION: (47)..(47)  
OTHER INFORMATION: "X" is defined as any amino acid residue.  
FEATURE:  
NAME/KEY: MISC\_FEATURE  
LOCATION: (51)..(52)  
OTHER INFORMATION: "X" is defined as any amino acid residue.

FEATURE:  
NAME/KEY: MISC\_FEATURE  
LOCATION: (58)..(59)  
OTHER INFORMATION: "X" is defined as any amino acid residue.  
FEATURE:  
NAME/KEY: MISC\_FEATURE  
LOCATION: (61)..(61)  
OTHER INFORMATION: "X" is defined as any amino acid residue.  
FEATURE:  
NAME/KEY: MISC\_FEATURE  
LOCATION: (73)..(73)  
OTHER INFORMATION: "X" is defined as any amino acid residue.  
FEATURE:  
NAME/KEY: MISC\_FEATURE  
LOCATION: (77)..(78)  
OTHER INFORMATION: "X" is defined as any amino acid residue.  
FEATURE:  
NAME/KEY: MISC\_FEATURE  
LOCATION: (86)..(87)  
OTHER INFORMATION: "X" is defined as any amino acid residue.  
FEATURE:  
NAME/KEY: MISC\_FEATURE  
LOCATION: (95)..(95)  
OTHER INFORMATION: "X" is defined as any amino acid residue.  
FEATURE:  
NAME/KEY: MISC\_FEATURE  
LOCATION: (97)..(97)  
OTHER INFORMATION: "X" is defined as any amino acid residue.  
FEATURE:  
NAME/KEY: MISC\_FEATURE  
LOCATION: (99)..(99)  
OTHER INFORMATION: "X" is defined as any amino acid residue.  
FEATURE:  
NAME/KEY: MISC\_FEATURE  
LOCATION: (102)..(102)  
OTHER INFORMATION: "X" is defined as any amino acid residue.  
FEATURE:  
NAME/KEY: MISC\_FEATURE  
LOCATION: (106)..(106)  
OTHER INFORMATION: "X" is defined as any amino acid residue.  
FEATURE:  
NAME/KEY: MISC\_FEATURE  
LOCATION: (110)..(110)  
OTHER INFORMATION: "X" is defined as any amino acid residue.  
FEATURE:  
NAME/KEY: MISC\_FEATURE  
LOCATION: (114)..(114)  
OTHER INFORMATION: "X" is defined as any amino acid residue.  
FEATURE:  
NAME/KEY: MISC\_FEATURE  
LOCATION: (117)..(118)  
OTHER INFORMATION: "X" is defined as any amino acid residue.  
FEATURE:  
NAME/KEY: MISC\_FEATURE  
LOCATION: (120)..(120)  
OTHER INFORMATION: "X" is defined as any amino acid residue.  
US-10-043-539-26

Query Match 39.0%; Score 224; DB 15; Length 120;  
Best Local Similarity 43.9%; Pred. No. 4.8e-15;  
Matches 50; Conservative 15; Mismatches 49; Indels 0; Gaps 0;  
Qy 1 MSKINDINLVNATFQVKXFFRDYKKFNLYNVEEYILNHLRSNLSISKEIAKCSF 60  
Db 3 MXKINDIXDLNXXYXXFKXIXKEFXLSFEEXFLTYIXXQXENEXXLDIIXXLY 62  
Qy 61 KPYLTKLOKLDKLLSKSKESLQDRTVIVVVTDTQKANIQKLISELEYIK 114  
Db 63 KQPLVKAUKLKKXXYLSKSKSXXDRTVLIXVXDQKRXKIELLSXVNOXIK 116

RESULT 3  
US-10-290-142-9

```
/ Sequence 9, Application US/10290142
/ Publication No. US20030100002A1
/ GENERAL INFORMATION:
/ APPLICANT: Cheung, Ambrose
/ TITLE OF INVENTION: METHODS FOR PREVENTING OR REDUCING MULTIPLE ANTIBIOTIC RESISTANCE
/ TITLE OF INVENTION: BACTERIA
/ FILE REFERENCE: DC-0195
/ CURRENT APPLICATION NUMBER: US/10/290,142
/ PRIOR FILING DATE: 2002-11-06
/ PRIOR APPLICATION NUMBER: US 10/092,264
/ PRIOR FILING DATE: 2002-03-06
/ PRIOR APPLICATION NUMBER: US 60/329,140
/ PRIOR FILING DATE: 2001-10-12
/ PRIOR APPLICATION NUMBER: US 60/312,546
/ PRIOR FILING DATE: 2001-08-15
/ PRIOR APPLICATION NUMBER: US 60/273,791
/ PRIOR FILING DATE: 2001-03-06
/ NUMBER OF SEQ ID NOS: 15
/ SOFTWARE: PatentIn version 3.1
/ SEQ ID NO 9
/ LENGTH: 124
/ TYPE: PRT
/ ORGANISM: Staphylococcus aureus
US-10-290-142-9

Query Match      25.8%; Score 148; DB 15; Length 124;
Best Local Similarity 30.1%; Pred. No. 2e-07;
Matches 34; Conservative 27; Mismatches 52; Indels 0; Gaps 0;

QY      1 MSKINDINDLVNATFQVKFFRDTKKKNLNVVEEIVILNHLRSSEISKEIAKCSF 60
      3 ITKINDCFELLSMWYADKLKSLIKKEFSISFEFAVLTYISENKEEYVLDIINHLY 62
DB      1 KPYVLTALQKLKDLKLSKRSLODERTVIVVYVDTQKANIQLISELEYI 113
      3 ITKINDCFELLSMWYADKLKSLIKKEFSISFEFAVLTYISENKEEYVLDIINHLY 62
QY      61 KPYVLTALQKLKDLKLSKRSLODERTVIVVYVDTQKANIQLISELEYI 113
      63 KQPQVVKAVKILSQEDYFDKGRNEHDETRVILVNAQQRKIESLILSRVKNRI 115
DB      1 MSKINDINDLVNATFQVKFFRDTKKKNLNVVEEIVILNHLRSSEISKEIAKCSF 60
      3 ITKINDCFELLSMWYADKLKSLIKKEFSISFEFAVLTYISENKEEYVLDIINHLY 62

RESULT 4
US-10-043-539-3
/ Sequence 3, Application US/10043539
/ Publication No. US20030114650A1
/ GENERAL INFORMATION:
/ APPLICANT: Manna, Adar
/ APPLICANT: Zhang, Gongyi
/ TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR AFFECTING VIRULENCE DETERMINANTS IN
/ TITLE OF INVENTION: BACTERIA
/ FILE REFERENCE: DC-0199
/ CURRENT APPLICATION NUMBER: US/10/043,539
/ CURRENT FILING DATE: 2002-01-11
/ PRIOR APPLICATION NUMBER: US 60/261,233
/ PRIOR FILING DATE: 2001-01-12
/ PRIOR APPLICATION NUMBER: US 60/261,607
/ PRIOR FILING DATE: 2001-01-12
/ PRIOR APPLICATION NUMBER: US 60/289,601
/ PRIOR FILING DATE: 2001-05-08
/ NUMBER OF SEQ ID NOS: 34
/ SOFTWARE: PatentIn version 3.1
/ SEQ ID NO 3
/ LENGTH: 124
/ TYPE: PRT
/ ORGANISM: Staphylococcus aureus
US-10-043-539-3

Query Match      25.8%; Score 148; DB 15; Length 124;
Best Local Similarity 30.1%; Pred. No. 2e-07;
Matches 34; Conservative 27; Mismatches 52; Indels 0; Gaps 0;

QY      1 MSKINDINDLVNATFQVKFFRDTKKKNLNVVEEIVILNHLRSSEISKEIAKCSF 60
      3 ITKINDCFELLSMWYADKLKSLIKKEFSISFEFAVLTYISENKEEYVLDIINHLY 62
DB      1 KPYVLTALQKLKDLKLSKRSLODERTVIVVYVDTQKANIQLISELEYI 113
      3 ITKINDCFELLSMWYADKLKSLIKKEFSISFEFAVLTYISENKEEYVLDIINHLY 62

RESULT 5
US-10-290-143-8
/ Sequence 8, Application US/10290143
/ Publication No. US20030124597A1
/ GENERAL INFORMATION:
/ APPLICANT: Cheung, Ambrose
/ TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR IDENTIFYING AGENTS WHICH REGULATE AU
/ TITLE OF INVENTION: PROCESSES IN BACTERIA
/ FILE REFERENCE: DC-0202
/ CURRENT APPLICATION NUMBER: US/10/290,143
/ CURRENT FILING DATE: 2002-11-06
/ PRIOR APPLICATION NUMBER: US 10/092,264
/ PRIOR FILING DATE: 2002-03-06
/ PRIOR APPLICATION NUMBER: US 60/329,140
/ PRIOR FILING DATE: 2001-10-12
/ PRIOR APPLICATION NUMBER: US 60/312,546
/ PRIOR FILING DATE: 2001-08-15
/ PRIOR APPLICATION NUMBER: US 60/273,791
/ PRIOR FILING DATE: 2001-03-06
/ NUMBER OF SEQ ID NOS: 17
/ SOFTWARE: PatentIn version 3.1
/ SEQ ID NO 8
/ LENGTH: 124
/ TYPE: PRT
/ ORGANISM: Staphylococcus aureus
US-10-290-143-8

Query Match      25.8%; Score 148; DB 15; Length 124;
Best Local Similarity 30.1%; Pred. No. 2e-07;
Matches 34; Conservative 27; Mismatches 52; Indels 0; Gaps 0;

QY      1 MSKINDINDLVNATFQVKFFRDTKKKNLNVVEEIVILNHLRSSEISKEIAKCSF 60
      3 ITKINDCFELLSMWYADKLKSLIKKEFSISFEFAVLTYISENKEEYVLDIINHLY 62
DB      1 KPYVLTALQKLKDLKLSKRSLODERTVIVVYVDTQKANIQLISELEYI 113
      63 KQPQVVKAVKILSQEDYFDKGRNEHDETRVILVNAQQRKIESLILSRVKNRI 115
DB      1 KPYVLTALQKLKDLKLSKRSLODERTVIVVYVDTQKANIQLISELEYI 113
      63 KQPQVVKAVKILSQEDYFDKGRNEHDETRVILVNAQQRKIESLILSRVKNRI 115

RESULT 6
US-10-145-602-4
/ Sequence 4, Application US/10145602
/ Publication No. US20030171563A1
/ GENERAL INFORMATION:
/ APPLICANT: McNamara, Peter
/ TITLE OF INVENTION: REGULATORS OF BACTERIAL VIRULENCE FACTOR EXPRESSION
/ FILE REFERENCE: KCC 4766
/ CURRENT APPLICATION NUMBER: US/10/145,602
/ CURRENT FILING DATE: 2002-05-13
/ PRIOR APPLICATION NUMBER: US 60/291,917
/ PRIOR FILING DATE: 2001-05-18
/ NUMBER OF SEQ ID NOS: 19
/ SOFTWARE: PatentIn version 3.1
/ SEQ ID NO 4
/ LENGTH: 247
/ TYPE: PRT
/ ORGANISM: Staphylococcus aureus
US-10-145-602-4

Query Match      21.4%; Score 123; DB 12; Length 247;
Best Local Similarity 27.8%; Pred. No. 0.00015;
Matches 30; Conservative 26; Mismatches 52; Indels 0; Gaps 0;

QY      7 INDLVNATFQVKFFRDTKKKNLNVVEEIVILNHLRSSEISKEIAKCSF 66
      9 VNRFINVKAVIFFLTQELKQYKLSKELLILAYFYFKNEHSLSLKEITGDIYKQSDVV 68
DB      1 KPYVLTALQKLKDLKLSKRSLODERTVIVVYVDTQKANIQLISELEYI 113
      63 KQPQVVKAVKILSQEDYFDKGRNEHDETRVILVNAQQRKIESLILSRVKNRI 115
```

QY 67 KALOKLKDLSKRSQDERTVIVVYDTQKANIQKLISELEYIK 114  
Db 69 KNKLSKKGFINRNEADRRIFSVTPIQRKIAACVINELDKIIR 116

## RESULT 7

US-10-145-602-2  
; Sequence 2, Application US/10145602  
; Publication No. US20030171563A1  
; GENERAL INFORMATION:  
; APPLICANT: McNamara, Peter  
; TITLE OF INVENTION: REGULATORS OF BACTERIAL VIRULENCE FACTOR EXPRESSION  
; FILE REFERENCE: KCC 4766  
; CURRENT APPLICATION NUMBER: US/10/145,602  
; CURRENT FILING DATE: 2002-05-13  
; PRIOR APPLICATION NUMBER: US 60/291,917  
; PRIOR FILING DATE: 2001-05-18  
; NUMBER OF SEQ ID NOS: 19  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 2  
; LENGTH: 166  
; TYPE: PRT  
; ORGANISM: Staphylococcus aureus  
US-10-145-602-2

Query Match 20.9%; Score 120; DB 12; Length 166;  
Best Local Similarity 28.0%; Pred. No. 0.00018;  
Matches 33; Conservative 26; Mismatches 45; Indels 14; Gaps 4;  
QY 1 MSKINDINDLVNATQ-----VKFFPRDTKKFNLYBEIYILNHLRSENEISKE 53  
Db 34 MKKNV--NDTVFQILQLETLGDSINFSIESEYKMSREEILIL--LTLWQKGFMTLKE 89  
QY 54 IAKCSEFKPYLTAKLOKLDLSKRSQDERTVIVVYDTQKANIQKLISE 108  
Db 90 MDRFVEVPYKRTIYNNLVELEIYKRPVDDERTVIIHFNEKLOQKVELLNFI 147

## RESULT 8

US-10-043-539-28  
; Sequence 28, Application US/10043539  
; Publication No. US20030114650A1  
; GENERAL INFORMATION:  
; APPLICANT: Cheung, Ambrose  
; APPLICANT: Manna, Adar  
; APPLICANT: Zhang, Gongyi  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR AFFECTING VIRULENCE DETERMINANTS IN  
; FILE REFERENCE: DC-0199  
; CURRENT APPLICATION NUMBER: US/10/043,539  
; CURRENT FILING DATE: 2002-01-11  
; PRIOR APPLICATION NUMBER: US 60/261,233  
; PRIOR FILING DATE: 2001-01-12  
; PRIOR APPLICATION NUMBER: US 60/261,607  
; PRIOR FILING DATE: 2001-01-12  
; PRIOR APPLICATION NUMBER: US 60/289,601  
; PRIOR FILING DATE: 2001-05-08  
; NUMBER OF SEQ ID NOS: 34  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 28  
; LENGTH: 118  
; TYPE: PRT  
; ORGANISM: Staphylococcus aureus  
US-10-043-539-28

Query Match 20.0%; Score 115; DB 15; Length 118;  
Best Local Similarity 25.0%; Pred. No. 0.00037;  
Matches 27; Conservative 26; Mismatches 55; Indels 0; Gaps 0;  
QY 6 DINDLVNATFQVKFFRDTKKFNLYBEIYILNHLRSENEISKEIAKCSFKPYLT 65  
Db 7 DSKEFLNLMYTMVFKNIKKHLLTSFVEFTILAITSQNKVILLKDLIETIHHKYPQT 66

QY 66 TKALOKLKDLSKRSQDERTVIVVYDTQKANIQKLISELEYIK 113  
Db 67 VRALNNLKFGYLIKERSTEDERKILIHMDDAQDHAEQLLAQVNNQLL 114

## RESULT 9

US-10-043-539-27  
; Sequence 27, Application US/10043539  
; Publication No. US20030114650A1  
; GENERAL INFORMATION:  
; APPLICANT: Cheung, Ambrose  
; APPLICANT: Manna, Adar  
; APPLICANT: Zhang, Gongyi  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR AFFECTING VIRULENCE DETERMINANTS IN  
; FILE REFERENCE: DC-0199  
; CURRENT APPLICATION NUMBER: US/10/043,539  
; CURRENT FILING DATE: 2002-01-11  
; PRIOR APPLICATION NUMBER: US 60/261,233  
; PRIOR FILING DATE: 2001-01-12  
; PRIOR APPLICATION NUMBER: US 60/261,607  
; PRIOR FILING DATE: 2001-01-12  
; PRIOR APPLICATION NUMBER: US 60/289,601  
; PRIOR FILING DATE: 2001-05-08  
; NUMBER OF SEQ ID NOS: 34  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 27  
; LENGTH: 119  
; TYPE: PRT  
; ORGANISM: Staphylococcus aureus  
US-10-043-539-27

Query Match 16.6%; Score 95; DB 15; Length 119;  
Best Local Similarity 21.3%; Pred. No. 0.038;  
Matches 23; Conservative 29; Mismatches 56; Indels 0; Gaps 0;  
QY 7 INDLVNATFQVKFFRDTKKFNLYBEIYILNHLRSENEISKEIAKCSFKPYLT 66  
Db 9 IRDFIIIEAYMPFRKKVKPEVDMTKEFILLTYLFHQQENTLPFKIVSDLCYKQSDLV 68  
QY 67 KALOKLKDLSKRSQDERTVIVVYDTQKANIQKLISELEYIK 114  
Db 69 QHIKVLVHSHYISKVRKIDERTYISISEQREKIAERVLFDQIIR 116

## RESULT 10

US-10-092-264-4  
; Sequence 4, Application US/10092264  
; Publication No. US20020164761A1  
; GENERAL INFORMATION:  
; APPLICANT: Trustees of Dartmouth College  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR REGULATING AUTOLYTIC PROCESSES IN BAC  
; FILE REFERENCE: DC-0188  
; CURRENT APPLICATION NUMBER: US/10/092,264  
; CURRENT FILING DATE: 2002-03-06  
; PRIOR APPLICATION NUMBER: US 60/273,791  
; PRIOR FILING DATE: 2001-03-06  
; PRIOR APPLICATION NUMBER: US 60/312,546  
; PRIOR FILING DATE: 2001-08-15  
; PRIOR APPLICATION NUMBER: US 60/329,140  
; PRIOR FILING DATE: 2001-10-12  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 4  
; LENGTH: 134  
; TYPE: PRT  
; ORGANISM: Staphylococcus aureus  
US-10-092-264-4

Query Match 16.3%; Score 93.5; DB 14; Length 134;  
Best Local Similarity 28.4%; Pred. No. 0.062;  
Matches 25; Conservative 25; Mismatches 35; Indels 3; Gaps 3;

```

Query Match      16.3%; Score 93.5; DB 14; Length 147;
Best Local Similarity 28.4%; Pred. No. 0.069;
Matches 25; Conservative 25; Mismatches 35; Indels 3; Gaps 3;

QY 10 LVNATFQVKKFRD-TKKKPNLYBEIYILNHLRSSEISKEIAKSEFKPYLTKA 68
      ||| : : : ||| : : : ||| : : : ||| : : :
Db 15 LVNAQRQVRYYASNKYFKYNTLPQFLVLT-ILWDES-PVNVKVVTEALDTGTVSPL 72
      ||| : : : ||| : : : ||| : : : ||| : : :

QY 69 LQKLKDLKLLSKRSIQDERTIVVYVTD 96
      ||| : : : ||| : : : ||| : : : ||| : : :
Db 73 LKMEQVDLLIKRSEVDQREVFHILTD 100
      ||| : : : ||| : : : ||| : : : ||| : : :

```

```

RESULT 13
US10-290-142-2
; Sequence 2, Application US/10290142
; Publication No. US20030100002A1
; GENERAL INFORMATION:
; APPLICANT: Cheung, Ambrose
; TITLE OF INVENTION: METHODS FOR PREVENTING OR REDUCING MULTIPLE ANTIBIOTIC RESISTANCE
; TITLE OF INVENTION: BACTERIA
; FILE REFERENCE: DC-0195

```

	Query Watch	16.3%; Score 93.5; DB 15; Length 147;
	Best Local Similarity	28.4%; Pred.No.0.069; Indels 3; Gaps 3;
	Matches	25; Conservative 25; Mismatches 35;
QY	10 LVNATPQVKFPRD-TKKFENLVYEIYLINHLRSNESISKETAKCEFPXYITKA	68
	: :: :     :     :     :     :     :     :	:: :: :: :: :: :: :: :: :: :: :: :: :: ::
Dd	15 LYNAAQVARNYSNKVFKNYLTYPQLVLTLWDES-FVNPKVVTEALDTGTVSPL	72
QY	69 LQKLDKLKLSKRSLQDERTVIYYVD	96
	: :: :     :     :     :     :     :	:: :: :: :: :: :: :: :: :: :: :: :: :: ::
Dd	73 LKRMEOVDLTKRSEVDQREVFILTD	100

```

RESULT 14
US-10-290-143-3
/ Sequence 3, Application US/10280143
/ Publication No. US20030124597A1
/ GENERAL INFORMATION:
/ APPLICANT: Cheung, Ambrose
/ TITLE OF INVENTION: COMPOSITIONS AND METHODS
/ TITLE OF INVENTION: PROCESSES IN BACTERIA
/ FILE REFERENCE: DC-0202
/ CURRENT APPLICATION NUMBER: US/10/290,143
/ CURRENT FILING DATE: 2002-11-06
/ PRIOR APPLICATION NUMBER: US 10/092,264
/ PRIOR FILING DATE: 2002-03-06
/ PRIOR APPLICATION NUMBER: US 60/329,140
/ PRIOR FILING DATE: 2001-10-12
/ PRIOR APPLICATION NUMBER: US 60/312,546
/ PRIOR FILING DATE: 2001-08-15
/ PRIOR APPLICATION NUMBER: US 60/273,791
/ PRIOR FILING DATE: 2001-03-06
/ NUMBER OF SEQ IDS NOS: 17

```

FILE NO.: CHEUNG, AMOL  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR IDENTIFYING AGENTS WHICH REGULATE AUTOPHAGY  
FILE REFERENCE: DC-0202  
DATE OF INVENTION: 2002-11-06  
CURRENT FILING DATE: 2002-11-06  
PRIORITY APPLICATION NUMBER: US 10/290,143  
PRIOR APPLICATION NUMBER: US 10/092,264  
PRIOR FILING DATE: 2002-03-06  
PRIOR APPLICATION NUMBER: US 60/329,140  
PRIOR FILING DATE: 2001-10-12  
PRIOR APPLICATION NUMBER: US 60/312,546  
PRIOR FILING DATE: 2001-08-15  
PRIOR APPLICATION NUMBER: US 60/273,791  
PRIOR FILING DATE: 2001-03-06  
NUMBER OF SEQ ID NOS: 17

```
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 147
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-10-290-143-3
```

```
Query Match      16.3%; Score 93.5; DB 15; Length 147;
Best Local Similarity 28.4%; Pred. No. 0.069;
Matches 25; Conservative 25; Mismatches 35; Indels 3; Gaps 3;

QY 10 LVNATFQVKKFRD-TKKFNLYEEIYLNHILRSSEISKEIAKCSFEKPYLTKA 68
   ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db 15 LYNARQVNRYSNKKVFKKLYTPQFLVLT-ILWDES-PVNVKVVTEALDGTGVSPL 72
   ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :

QY 69 LQKLDKLLSKRSLODERTVIVVTD 96
   ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db 73 LKRMQVDLIKRESEVDQREVFIHLTD 100
   ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
```

```
RESULT 15
US-10-290-142-8
; Sequence 8, Application US/10290142
; Publication No. US2003010002A1
; GENERAL INFORMATION:
; APPLICANT: Cheung, Ambrose
; TITLE OF INVENTION: METHODS FOR PREVENTING OR REDUCING MULTIPLE ANTIBIOTIC RESISTANCE
; FILE REFERENCE: DC-0195
; CURRENT APPLICATION NUMBER: US/10/290,142
; CURRENT FILING DATE: 2002-11-06
; PRIOR APPLICATION NUMBER: US 10/092,264
; PRIOR FILING DATE: 2002-03-06
; PRIOR APPLICATION NUMBER: US 60/329,140
; PRIOR FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/312,546
; PRIOR FILING DATE: 2001-08-15
; PRIOR APPLICATION NUMBER: US 60/273,791
; PRIOR FILING DATE: 2001-03-06
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 8
; LENGTH: 153
; TYPE: PRT
; ORGANISM: Xanthomonas campestris
US-10-290-142-8
```

```
Query Match      15.9%; Score 91; DB 15; Length 153;
Best Local Similarity 26.4%; Pred. No. 0.13;
Matches 24; Conservative 21; Mismatches 44; Indels 2; Gaps 1;

QY 10 LVNATFQVKKFRD-TKKFNLYEEIYLNHILRSSEISKEIAKCSFEKPYLTKA 69
   ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db 25 LYSANLAKLYRGLLTKTLDLTPCYLVVLWENDGRSVS--EIGERYLDSATLTPLL 82
   ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :

QY 70 QKLDKLLSKRSLODERTVIVVTDQKA 100
   ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db 83 KELESAGLLTRAAHDERQVIIGIADAGRA 113
   ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
```

Search completed: November 17, 2003, 12:54:10  
Job time : 267 secs



C:Genetics:  
A:Gene: sarA

Query Match 25.8%; Score 148; DB 2; Length 124;  
Best Local Similarity 30.1%; Pred. No. 2.5e-05;  
Matches 34; Conservative 27; Mismatches 5; Indels 0; Gaps 0;  
QY 1 MSKINDINDLVNATQVKKFFRDTKKFNLYEEIYILNHILRSSENEISSEIKIAKCEFPYLYT 60  
DB 3 ITKINDCFELLSMTVYADKLSLIKKEFSISFEPAVLTYISENKEEYVYKLDIINHLNY 62

QY 61 KPYLYTKALQKLDKLSKRSLODERTVIVYDTOKANIQKLISELEYI 113  
DB 63 KQPVVKAVKILSQBEDYDKRNEHDERTVILVNAQQRKKIESLLGRVNRKI 115

## RESULT 3

F89961  
repressor of toxins Rot [imported] - Staphylococcus aureus (strain N315)  
C:Species: Staphylococcus aureus  
C>Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 22-Oct-2001  
C:Accession: F89961  
R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Oguchi, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.; C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.  
Lancet 357, 1225-1240, 2001  
A:Title: Whole genome sequencing of meticillin-resistant Staphylococcus aureus.  
A:Reference number: A89758; PMID:21311952; PMID:11418146  
A:Accession: F89961  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-153 <KUR>  
A:Cross-references: GB:BA000018; PID:gl3701558; PIDN:BA842851.1; GSPDB:GN00149  
A:Experimental source: strain N315  
C:Genetics:  
A:Gene: rot

Query Match 21.4%; Score 123; DB 2; Length 153;  
Best Local Similarity 28.0%; Pred. No. 0.0033;  
Matches 33; Conservative 26; Mismatches 45; Indels 14; Gaps 4;  
QY 1 MSKINDINDLVNATQ-----VKKFFRDTKKFNLYEEIYILNHILRSSENEISSE 53  
DB 21 MKVAN--NDTVPGILQLETLGDINSIFSEIESEYKMSREBIL--LTLMQKSGMTLKE 76  
QY 54 IAKCSEFPYLYTKALQKLDKLSKRSLODERTVIVYDT--TKANIQKLI 108  
DB 77 MDRFVEVPYKTRTYNNLVEWLYKRPVDDERTVHHNEKLOQKVELLNFI 134

## RESULT 4

D90053  
hypothetical protein sarH2 [imported] - Staphylococcus aureus (strain N315)  
C:Species: Staphylococcus aureus  
C>Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 22-Oct-2001  
C:Accession: D90053  
R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Oguchi, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.; C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.  
Lancet 357, 1225-1240, 2001  
A:Title: Whole genome sequencing of meticillin-resistant Staphylococcus aureus.  
A:Reference number: A89758; PMID:21311952; PMID:11418146  
A:Accession: D90053  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-247 <KUR>  
A:Cross-references: GB:BA000018; PID:gl3702449; PIDN:BA843590.1; GSPDB:GN00149  
A:Experimental source: strain N315  
C:Genetics:  
A:Gene: sarH2

Query Match 21.4%; Score 123; DB 2; Length 247;  
Best Local Similarity 27.8%; Pred. No. 0.0056;

Matches 30; Conservative 26; Mismatches 52; Indels 0; Gaps 0;  
QY 7 INDLVNATQVKKFFRDTKKFNLYEEIYILNHILRSSENEISSEIKIAKCEFPYLYT 66  
DB 9 VNKFINVEAYIFFLQELKQYKLSIKELLILAYFYKNEHSISLKEIIGDILYKQSDVW 68  
QY 67 KALQKLDKLSKRSLODERTVIVYDTOKANIQKLISELEYI 114  
DB 69 KNIKSLKSGFINKRNEADERIFVSVTPIQKKIACVINELDKIK 116

## RESULT 5

D89771  
hypothetical protein sarH1 [imported] - Staphylococcus aureus (strain N315)  
C:Species: Staphylococcus aureus  
C>Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 22-Oct-2001  
C:Accession: D89771  
R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Oguchi, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.; C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.  
Lancet 357, 1225-1240, 2001  
A:Title: Whole genome sequencing of meticillin-resistant Staphylococcus aureus.  
A:Reference number: A89758; PMID:21311952; PMID:11418146  
A:Accession: D89771  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-250 <KUR>  
A:Cross-references: GB:BA000018; PID:gl3700028; PIDN:BA841327.1; GSPDB:GN00149  
A:Experimental source: strain N315  
C:Genetics:  
A:Gene: sarH1

Query Match 20.0%; Score 115; DB 2; Length 250;  
Best Local Similarity 25.0%; Pred. No. 0.025;  
Matches 27; Conservative 26; Mismatches 55; Indels 0; Gaps 0;  
QY 6 DINDLVNATQVKKFFRDTKKFNLYEEIYILNHILRSSENEISSEIKIAKCEFPYLYT 65  
DB 132 DSKEFLNLMWTFYFNKIIKKHLLTSLFVEFTILAITSQNKVILKOLLIETIRHKYPQT 191  
QY 66 TYALQKLDKLSKRSLODERTVIVYDTOKANIQKLISELEYI 113  
DB 192 VFLANNLKQGYLIKERSTEDERKILIHMDDAQDQHAFLQAQVNL 239

## RESULT 6

F90024  
hypothetical protein [imported] - Staphylococcus aureus (strain N315)  
C:Species: Staphylococcus aureus  
C>Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 22-Oct-2001  
C:Accession: F90024  
R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Oguchi, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.; C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.  
Lancet 357, 1225-1240, 2001  
A:Title: Whole genome sequencing of meticillin-resistant Staphylococcus aureus.  
A:Reference number: A89758; PMID:21311952; PMID:11418146  
A:Accession: F90024  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-116 <KUR>  
A:Cross-references: GB:BA000018; PID:gl3702067; PIDN:BA843359.1; GSPDB:GN00149  
A:Experimental source: strain N315  
C:Genetics:  
A:Gene: SA2062

Query Match 17.7%; Score 101.5; DB 2; Length 116;  
Best Local Similarity 29.3%; Pred. No. 0.13;  
Matches 29; Conservative 23; Mismatches 38; Indels 9; Gaps 3;

QY 16 QVKKFFRDTKKFNLYEEIYILNHILRSSENEISSEIKIAKCEFPYLYT-----AQK 71  
DB 17 QLKHLKTTTHK---ISIEEFVVLVKY--EAEKISGKELRDLTHFEMLDTSKIDVIIRK 71



Dp 37 DLSIEOFFVLREI--ATGDEISATELSACLNVNKSAITPKLKLFEKGYIRPERNKHDKR 94

Dp 37 DLSIEOFFVLREI--ATGDEISATELSACLNVNKSAITPKLKLFEKGYIRPERNKHDKR 94

QY : 69 LQKLKDLKLLSKKRSIQDERTVIVVTD 96

QY : 69 LQKLKDLKLLSKKRSIQDERTVIVVTD 96

```
Db 73 LKRMQVDLIKRESEVQDREVFHILTD 100

RESULT 11
G97236
A:Prase involved in DNA repair [imported] - Clostridium acetobutylicum
C:Species: Clostridium acetobutylicum
C>Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 14-Sep-2001
C:Accession: G97236
R:Noelling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,
J.; Daly, W.G.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
J. Bacteriol. 183, 4823-4838, 2001
A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clostridium acetobutylicum
A:Accession: G97236
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1163 <KUR>
A:Cross-references: GB:BA001437; PIDN:AAK90682.1; PID:gl5025772; GSPDB:GN00168
A:Experimental source: Clostridium acetobutylicum ATCC824
C:Genetics:
A:Gene: CAC2736

Query Match 16.2%; Score 93; DB 2; Length 1163;
Best Local Similarity 29.8%; Pred. No. 8;
Matches 36; Conservative 18; Mismatches 39; Indels 28; Gaps 5;

QY 3 KINDINDLVNATFQVKKFFRDTKKFN-LNYEEIIVLNHILRSSENEISKEIAKCSF 61
Db 873 KVNDI-DKLNKILMEKEFENIEKAKENYLNKDSVDEKYKNSLVNGA----- 925

QY 62 PYYLTKALQKLDKLSKK---RSLQDERTVIVVDTQKANTQKLISE-----LEBYIK 114
Db 926 -----VEVLSKLNKRLTEKWIETQNNRVERAKAKALQERSIKLEBEVK 972

QY 115 N 115
Db 973 N 973

RESULT 12
D90028
A:Hypothetical protein [imported] - Staphylococcus aureus (strain N315)
C:Species: Staphylococcus aureus
C>Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 22-Oct-2001
C:Accession: D90028
R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Ogud
ma, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.;
C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.
Lancet 357, 1225-1240, 2001
A:Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.
A:Reference number: A9759; MUID:21311952; PMID:11418146
A:Accession: D90028
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-247 <KUR>
A:Cross-references: GB:BA000018; PID:gl3702097; PIDN:BA043389.1; GSPDB:GN00149
A:Experimental source: strain N315
C:Genetics:
A:Gene: SA2091

Query Match 16.0%; Score 92; DB 2; Length 247;
Best Local Similarity 29.7%; Pred. No. 1.8;
Matches 33; Conservative 20; Mismatches 50; Indels 8; Gaps 5;

QY 9 DLVNAFQVKKFFRDTKKFN---LNYEEIIVLNHILRSSENEISKEIAKCSF-KPY 64
Db 130 DVSYSLLSQELFNKCYLNCQTLBELVYG-ILNLHGQDVTXELQ--GEFHPIF 186

QY 65 LTKALQKLDK-LLSKRSQDERTVIVVDTQKANTQKLISELEBYIK 114
Db 187 AVSPILKCLIKGLVKERCDELDERVITIKREKFSKVTWLIQACYNLYE 237

Db 126 AVSPILKCLIKGLVKERCDELDERVITIKREKFSKVTWLIQACYNLYE 237

RESULT 13
T52550
A:Probable transcription regulator 16.1K [imported] - Bacillus firmus
C:Species: Bacillus firmus
C>Date: 24-Oct-2000 #sequence_revision 24-Oct-2000 #text_change 01-Mar-2002
C:Accession: T52550
R:Ito, M.; Cooperberg, B.; Krulwich, T.A.
submitted to the EMBL Data Library, February 1997
A:Description: Cloning and sequencing of Na+/H+ antiporter homolog gene from alkaliphilic
A:Reference number: Z26111
A:Accession: T52550
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-148 <ITO>
A:Cross-references: EMBL:U89914; PIDN:AAB87745.1
A:Experimental source: strain OP4
C:Superfamily: transcription regulator yfiV

Query Match 15.8%; Score 90.5; DB 2; Length 148;
Best Local Similarity 26.5%; Pred. No. 1.4;
Matches 26; Conservative 24; Mismatches 35; Indels 13; Gaps 3;

QY 25 KKKFNLYEEIIVLNHILRSSENEISKEIAKCSFDPVYLTAKQLKLDKLSKSL 84
Db 31 KLEHNLGTGEFFVLNTL--EQGRITSSQLABELQVKPSAITAMVDRLKKNDFVIREDE 88

QY 85 QDERTVIVVDTQKANTQKLISELEEE 111
Db 89 KORRAVYVVRISDEGRALKSSVKKNITMEKYNKLT 126

RESULT 14
G70163
A:Hypothetical protein B505.2 - Lyme disease spirochete
C:Species: Borrelia burgdorferi (Lyme disease spirochete)
C>Date: 13-Feb-1998 #sequence_revision 13-Feb-1998 #text_change 08-Oct-1999
C:Accession: G70163
R:Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; White,
son, D.; Peterson, J.; Kerlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vugt,
; Bowman, C.; Garland, S.; Fujil, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.
Nature 390, 580-586, 1997
A:Authors: Smith, H.O.; Venter, J.C.
A:Title: Genomic sequence of a Lyme disease spirochete, Borrelia burgdorferi.
A:Reference number: A70100; MUID:98065943; PMID:9403685
A:Accession: G70163
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-2166 <KLE>
A:Cross-references: GB:AE001153; GB:AE000783; PID:G2688419; PIDN:AAC66876.1; PID:G2688426
A:Experimental source: strain B31

Query Match 15.7%; Score 90; DB 2; Length 2166;
Best Local Similarity 32.2%; Pred. No. 27;
Matches 39; Conservative 18; Mismatches 44; Indels 20; Gaps 6;

QY 9 DLVNAFQVKKFFRDTKKFNLYEEIIVLNHILRSSENEISK---EIAKSEFPYYL 65
Db 997 DITLDSLVNK--FNDINKINGKYNV-ISNV--RGYSENISKLENEIMHEIENLSRRL 1051

QY 66 T-----KALQKLDKLSKRSQDERTVIVVDTQKANTQKLISELEBYIK 114
Db 1052 TDRISLSKGMENLQKLESFVSKYQ-VEKFELKVDLTDDGEAKINKLVKEIEQYIK 1110

QY 115 N 115
Db 1111 S 1111

RESULT 15
AC1556
A:regulatory protein (MarR family) homolog lin0988 [imported] - Listeria innocua (strain C)
C:Species: Listeria innocua
```

Search completed: November 17, 2003, 12:45:08  
Job time : 28 secs

**This Page Blank (uspto)**

```

Query Match      16.2%; Score 93; DB 1; Length 1163;
Best Local Similarity 29.8%; Pred. No. 4.2;
Matches 36; Conservative 18; Mismatches 39; Indels 28; Gaps 5;
QV 3 KINDINDLVNATFOVKFFPRDTRKFRN-LNVEEYILNHLHRSNEISGKTIACKSEPK 61

```

Db 873 KVNDI-DKLNKILMBEKPENTEKAKENYNDKNEINLLKSDVEKYKNELSKVNGA----- 925

QY 62 PYLTKAKQKLDKLLSKK---RSLQDERTVIVVVTDTOKANTOKLISE-----LEBYIK 114

Db 926 -----VEVLSKKLKNRKLTEKEWIEQNNRVEKAKAKALQERSIKLEBEVK 972

QY 115 N 115

Db 973 N 973

RESULT 2

ID\_ZIPI\_YEAST STANDARD; PRT; 875 AA.

AC P31111;

DT 01-JUL-1993 (Rel. 26, Created)

DT 01-OCT-1996 (Rel. 34, Last sequence update)

DT 15-JUL-1998 (Rel. 36, Last annotation update)

DE Synaptonemal complex protein ZIPI.

GN ZIPI OR YDR285W OR P9819.9.

OS Saccharomyces cerevisiae (Baker's yeast).

OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.

OX NCBI\_TaxID=4932;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=BR1924-3B;

RX MEDLINE=93161412; PubMed=7916652;

RA Sym M., Engelbrecht J.A., Roeder G.S.;

RT "ZIPI is a synaptonemal complex protein required for meiotic chromosome synapsis.";

RT Cell 72:365-378(1993).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=S288c / AB972;

RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Du Z., Favell A., Fulton L., Gattung S., Greco T., Kirsten J., Kucaba T., Hallsworth K., Hawkins J., Hillier L., Jier M., Johnson D., Johnston L., Langston Y., Latreille P., Le T., Mardis E., Menezes S., Miller N., Nhan M., Pauley A., Peluso D., Rifkin L., Riles L., Taich A., Trevasakis E., Vignati D., Wilcox L., Wohlman P., Vaudin M., Wilson R., Waterston R.; Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.

RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.

CC -!- FUNCTION: REQUIRED FOR MEIOTIC CHROMOSOME SYNAPSIS AND CELL CYCLE PROGRESSION. MAY ACT AS A MOLECULAR ZIPPER TO BRING HOMOLOGOUS CHROMOSOMES IN CLOSE APPPOSITION. ZIPI MAY ENCODE THE TRANSVERSE FILAMENTS OF THE SYNAPTONEMAL COMPLEX.

CC -!- SUBCELLULAR LOCATION: SYNAPSED MEIOTIC CHROMOSOMES.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; L06487; AAA35239.1; -.

DR EMBL; U51031; AA964474.1; -.

DR FIR; S70115; ZIPI.

DR SGD; S0002693; ZIPI.

DR GO; GO:0005716; C:synaptonemal complex; IDA.

DR GO; GO:0007126; P:meiosis; IMP.

DR GO; GO:0007129; P:synapsis; IMP.

KW Nuclear protein; Meiosis; Coiled coil.

FT DOMAIN 177 333 COILED COIL (POTENTIAL).

FT DOMAIN 357 438 COILED COIL (POTENTIAL).

FT DOMAIN 456 752 COILED COIL (POTENTIAL).

FT CONFLICT 55 T -> A (IN REF. 1).

FT SEQUENCE 875 AA; 100035 MW; 674F12625CD9DDFD CRC64;

Query Match 15.4%; Score 88.5; DB 1; Length 875;

Best Local Similarity 25.2%; Pred. No. 7;

Matches 32; Conservative 26; Mismatches 46; Indels 23; Gaps 4;

QY 8 NDIVNATFOVKFFRDTKKKFNLYVEEYILNHLRSNESISSEIKAKSCFEKPYLTK 67

Db 621 NQKQHTTKLEAFQKNNQOLQKLNVEVOLKAHELEEE---QNRHLKNCLEKKTGVEE 677

QY 68 ALQKLKDLK-----LLSKRSIQDERTVIVVVT-----DTQRANIQKLISE 108

Db 678 SLSDVTKLQQVILKSEKQDITAEKLEODNLESLEEVTKNLQKQVQSKRELEQIKE 737

QY 109 LEEYIKN 115

Db 738 LEE-IXN 743

RESULT 3

EXSB\_BORBU STANDARD; PRT; 1169 AA.

ID\_EXSB\_BORBU

AC OS1578;

DT 30-MAY-2000 (Rel. 39, Created)

DT 30-MAY-2000 (Rel. 39, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Exodeoxyribonuclease V beta chain (EC 3.1.11.5).

GN RECB OR BB0633.

OS Borrelia burgdorferi (Lyme disease spirochete).

OC Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Borrelia.

OX NCBI\_TaxID=139;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=ATCC 35210 / B31;

RX MEDLINE=98065943; PubMed=9403685;

RA Fraser C.M., Casjens S., Huang W.M., Sutton G.G., Clayton R.A., Lathigra R., White O., Ketchum K.A., Dodson R., Hickey E.K., Gwinn M., Peterson J., Tomb J.-F., Fleischmann R.D., Richardson D., Petersen J., Kerlavage A.R., Quackenbush J., Salzberg S., Hanson M., van Vugt R., Palmer N., Adams M.D., Gocayne J.D., Weidman J., Utterback T., Wathley L., McDonald L., Attiach P., Bowman C., Garland S., Fujii C., Cotton M.D., Horst K., Roberts K., Hatch B., Smith H.O., Venter J.C.;

RT "Genomic sequence of a Lyme disease spirochete, Borrelia burgdorferi.";

RL Nature 390:580-586(1997).

CC -!- FUNCTION: REQUIRED FOR EFFICIENT DNA REPAIR; IT CATALYZES THE UNWINDING OF DOUBLE-STRANDED DNA AND THE CLEAVAGE OF SINGLE-STRANDED DNA AND IT STIMULATES LOCAL GENETIC RECOMBINATION. ALL OF THESE ACTIVITIES REQUIRE CONCOMITANT HYDROLYSIS OF ATP (BY SIMILARITY).

CC -!- CATALYTIC ACTIVITY: Exonucleolytic cleavage (in the presence of ATP) in either 5'- to 3'- or 3'- to 5'-direction to yield 5'-phosphooligonucleotides.

CC -!- SUBUNIT: CONSIST OF THREE SUBUNITS; RECB, RECC AND RECD (BY SIMILARITY).

CC -!- SIMILARITY: BELONGS TO THE HELICASE FAMILY. UNWRD SUBFAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; AE001164; AAC66981.1; -.

DR EMBL; H70178; H70178.

DR HSSP; P56255; 1PUR.

DR TIGR; B30633; -.

DR InterPro; IPR004596; RecB.

DR InterPro; IPR000212; UvrD-helicase.

DR Pfam; PF00380; UvrD-Helicase; 1.

DR TIGRFAMs; TIGR00609; recB; 1.

DR Hydrolase; Nuclease; Exonuclease; Endonuclease; Helicase; ATP-binding; DNA repair; Complete proteome.

KW NP\_BIND 18 25 ATP (POTENTIAL).

FT NP\_BIND 18 25 ATP (POTENTIAL).

```

Query Match          14.7%; Score 84.5; DB 1; Length 301;
Best Local Similarity 26.11%; Pred. No. 4.6;
Matches 23; Conservative 23; Mismatches 27; Indels 15; Gaps 3;

QY 42 LRSENEISSKEITA-----KCFSEKPYV-----LTKALQKLKDLKLLSKRSIQDERTV 90
DB 17 LSRQNGIQTNPALHATQCSEFTLMARRIGKDLNNTFAKLEKILAKRSKLSLFDKAV 76
QY 91 ----IYVTDQKANTQKLISLEBYIK 114
DB 77 EIEELTYIIKQDINSLNKQIAQLQDFVR 104

RESULT 5
DYHB_HUMAN
ID QY6DT5; STANDARD; PRT; 4523 AA.
AC Q96DT5; Q9UJ82;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Ciliary dynein heavy chain 11 (Axonemal beta dynein heavy chain 11).
GN DNAH11.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID:9606;
RN [1]
SEQUENCE FROM N.A., VARIANT PCD GLN-3004, AND VARIANTS LEU-34;
ARG-639; CYS-654; ALA-1023; THR-1038; GLY-1640; ASN-2641; THR-3474;
VAL-3715; PRO-3765 AND ILE-4177.
TISSUE=Nasal epithelium, and Testis;
MEDLINE=22155903; PubMed=12142464;
RA Bartoloni L., Blouin J.L., Pan Y., Gehrig C., Maiti A.K., Scamuffa N.,
RA Rossier C., Jorissen M., Arnenget M., Meeks M., Mitchison E.M.,
RA Chung E.M., Delozier-Blanchet C.D., Craigen W.J., Antonarakis S.E.;
RA "Mutations in the DNAH11 (Axonemal heavy chain dynein type 11) gene
RT cause one form of situs inversus totalis and most likely primary
RT ciliary dyskinesia."
RT Proc. Natl. Acad. Sci. U.S.A. 99:10282-10286(2002).
RN [2]
SEQUENCE OF 1904-2004 FROM N.A.
RC TISSUE=Nasal polyps;
RC Maiti A.K., Mattei M.-G., Jorissen M., Voiz A., Ziegler A.,
RC Bouvagnet P.;
RA "Chromosomal localization of human dynein heavy chain genes."
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases
RL
CC -!- FUNCTION: Force generating protein of respiratory cilia. Produces
CC force towards the minus ends of microtubules. Dynein has ATPase
CC activity; the force-producing power stroke is thought to occur on
CC release of ADP.
CC
CC -!- SUBUNIT: Consists of at least two heavy chains and a number of
CC intermediate and light chains.
CC
CC -!- DOMAIN: Dynein heavy chains probably consist of an N-terminal stem
CC (which binds cargo and interacts with other dynein components),
CC and the head or motor domain. The motor contains six tandemly-
CC linked AAA domains in the head, which form a ring. A stalk-like
CC structure (formed by two of the coiled coil domains) protrudes
CC between AAA 4 and AAA 5 and terminates in a microtubule-binding
CC site. A seventh domain may also contribute to this ring; it is not
CC clear whether the N-terminus or the C-terminus forms this extra
CC domain. There are four well-conserved and two non-conserved ATPase
CC sites, one per AAA domain. Probably only one of these (within AAA
CC 1) actually hydrolyzes ATP, the others may serve a regulatory
CC function.
CC
CC -!- DISEASE: Defects in DNAH11 are a cause of immotile cilia syndrome
CC 1 (ICSI), also known as primary ciliary dyskinesia (PCD). The
CC phenotype of this autosomal recessive disease is characterized by
CC axonemal abnormalities of respiratory cilia and sperm tails
CC leading to bronchiectasis and sinusitis, which are sometimes
CC associated with situs inversus (Kartagener syndrome) and male
CC sterility.
CC
CC -!- SIMILARITY: Belongs to the dynein heavy chain family.

```

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

```

CC  EMBL; AJ320497; CAC60121.1; -
CC  EMBL; AJ132087; CAA10560.1; -
CC  Genew; HGNC:2942; DNAB11.
CC  MIM; 603339; -
CC  MIM; 242650; -
CC  InterPro; IPR003593; AAA ATPase.
CC  InterPro; IPR005613; AIP3.
CC  InterPro; IPR004273; Dynein_heavy.
CC  Pfam; PF03028; Dynein_heavy; 1.
CC  SMART; SM00382; AAA; 4.
CC  Motor protein; Dynein; Microtubules; ATP-binding; Repeat; Coiled coil;
KW  Polymorphism; Disease mutation.
FT  DOMAIN 1 1861 STEM (BY SIMILARITY).
FT  DOMAIN 1862 2083 AAA 1 (BY SIMILARITY).
FT  DOMAIN 2143 2373 AAA 2 (BY SIMILARITY).
FT  DOMAIN 2479 2726 AAA 3 (BY SIMILARITY).
FT  DOMAIN 2824 3073 AAA 4 (BY SIMILARITY).
FT  DOMAIN 3079 3410 STALK (BY SIMILARITY).
FT  DOMAIN 3466 3693 AAA 5 (BY SIMILARITY).
FT  DOMAIN 3903 4129 AAA 6 (BY SIMILARITY).
FT  DOMAIN 1274 1327 COILED COIL (POTENTIAL).
FT  DOMAIN 3079 3143 COILED COIL (POTENTIAL).
FT  DOMAIN 3319 3410 COILED COIL (POTENTIAL).
FT  DOMAIN 3675 3710 COILED COIL (POTENTIAL).
FT  NP_BIND 1900 1907 ATP (POTENTIAL).
FT  NP_BIND 2181 2188 ATP (POTENTIAL).
FT  NP_BIND 2517 2524 ATP (POTENTIAL).
FT  NP_BIND 2862 2869 ATP (POTENTIAL).
FT  VARIANT 34 34 E -> L (REQUIRES 2 NUCLEOTIDE
FT  SUBSTITUTIONS).
FT  VARIANT 639 639 /FTID=VAR_013851.
FT  VARIANT 654 654 Q -> R.
FT  VARIANT 1023 1023 /FTID=VAR_013852.
FT  VARIANT 1038 1038 S -> C.
FT  VARIANT 1640 1640 V -> A.
FT  VARIANT 2641 2641 /FTID=VAR_013853.
FT  VARIANT 2682 2682 /FTID=VAR_013854.
FT  VARIANT 3004 3004 A -> T.
FT  VARIANT 3474 3474 /FTID=VAR_013855.
FT  VARIANT 3715 3715 D -> G.
FT  VARIANT 3765 3765 S -> N.
FT  VARIANT 4177 4177 /FTID=VAR_013856.
FT  VARIANT 4523 4523 I -> V.
FT  VARIANT 520963 520963 /FTID=VAR_013857.
FT  VARIANT 7C9A71C95B296B89 7C9A71C95B296B89 /FTID=VAR_013858.
FT  VARIANT 14.5%; Score 83.5; DB 1; Length 4523;
FT  Best Local Similarity 26.9%; Pred. No. 1e+02;
FT  Matches 32; Conservative 24; Mismatches 38; Indels 25; Gaps 6;

```

```

Query Match 14.5%; Score 83.5; DB 1; Length 4523;
Best Local Similarity 26.9%; Pred. No. 1e+02;
Matches 32; Conservative 24; Mismatches 38; Indels 25; Gaps 6;

```

```

CC  7 INDVFNATFQVKFRDTPKFKFNLYEIIYILNHLRSENEISSEKSEFKPYLT 66
CC  VNEMSTRYQYRHHYTPKSFLEQISLFFKNLLKKQNEVSEKK-----ERLV 3107

```

```

QY 67 KALOKLK-----DLK--LLSKRSLO-----DERTVIVV--TDTOKANIKLISELE 111
DB 3108 NGIQKUKTTASQVGLKARLASQAEALQURNHDAALITKIGLQTEKVSREKTIADAE 3166

RESULT 6
DPOL CBEPV STANDARD; PRT; 964 AA.
AC P30319; 1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE DNA polymerase (EC 2.7.7.7).
OS Choriostoneura biennis entomopoxvirus (CbEPV).
OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Entomopoxvirinae;
OC Entomopoxvirus B.
CX NCBI_TaxID=10288;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92199242; PubMed=1840516;
RA Mustafa A., Yuen L.;
RT "Identification and sequencing of the Choriostoneura biennis
entomopoxvirus DNA polymerase gene.";
RL DNA Seq. 2:39-45(1991).
CC -!- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate = N diphosphate
CC + {DNA}(N).
CC -!- SIMILARITY: BELONGS TO THE DNA POLYMERASE TYPE-B FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announcement/
or send an email to license@isb-sib.ch).


EMBL; X57314; CAA40566.1; -
DR PIR; S25855; S25855.
DR InterPro; IPR006172; DNA_pol_B.
DR InterPro; IPR006134; DNA_pol_B_dom.
DR InterPro; IPR006133; DNA_pol_B_exo.
DR Pfam; PF00136; DNA_pol_B_1.
DR Pfam; PF03104; DNA_pol_B_exo; 1.
DR SMART; SM00486; POLSC; 1.
DR PROSITE; PS00116; DNA_POLYMERASE_B; 1.
DR Transferase; DNA-directed DNA polymerase; DNA replication;
KW DNA-binding.
SQ SEQUENCE 964 AA; 114818 MW; F8FF5AC957363A68 CRC64;

Query Match 14.4%; Score 82.5; DB 1; Length 964;
Best Local Similarity 23.2%; Pred. No. 23;
Matches 29; Conservative 23; Mismatches 40; Indels 33; Gaps 5;

QY 2 SKINDINDVFNATFQVKXF-----PRDTK-----KKFNLYEIIYILNHLRSENEISSEK 52
DB 765 NKFDNKEDLVNKSHECFQFLSNILNKNFPEYKFWIMWLLA----- 815

QY 53 ETAKSEFKPYLTALOKLKDCLKSKRSLODERTVIV--VTDTOKANIKLISELE 110
DB 816 -----KKYIGEVWSSMNPQLISDQSG-----TALIRDCTEHHTILKNTIDILK 862

QY 111 EYIKN 115
DB 863 EYLTN 867

RESULT 7
RASO METJA STANDARD; PRT; 1005 AA.
ID RASO METJA
AC Q58718;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)


```







Qy 4 INDINDLV-----NATFQVKKFFRDTKKKFNLNBYBEIYILNHLRSEISSEISS 52

Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,  
Pardo V., Pohl T.M., Portetelle D., Rapoport G., Prescott A.M.,  
Presecan E., Pujic P., Purnelle B., Rappott G., Rey M., Reynolds S.,  
Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadale Y.,  
Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,  
Sekiguchi J.J., Sekowska A., Seror S.J., Serroir P., Shin B.S., Soldo B.,  
Sorokin A., Tacconi E., Takagi I., Takahashi H., Takemaru K.,  
Takeuchi M., Tanakoshi A., Tanaka T., Terpstra P., Tognoni A.,  
Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,  
Viari A., Wambuit R., Wedler E., Wedler H., Weitzenecker T.,  
Winters P., Wibat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,  
Yoshida K., Yoshikawa H.F., Zumstein E., Yoshioka M., Zanchin A.,  
"The complete genome sequence of the Gram-positive bacterium Bacillus  
subtilis";,  
RT Nature 390:249-256 (1997).  
RT [3]  
RN FUNCTION.  
RP STRAIN=168;  
RC MEDLINE=21311737; PubMed=11418552;  
RA Puangthong M., Atichartpongkul S., Mongkolsuk S., Helmann J.D.;  
RT "OhrR is a repressor of OhrA, a key organic hydroperoxide resistance  
RT determinant in Bacillus subtilis";,  
RL J. Bacteriol. 183:4134-4141 (2001).  
RT [4]  
RN  
RX OXIDATION OF CY5-15, AND MUTAGENESIS OF CY5-15.  
RP MEDLINE=22008016; PubMed=11983871;  
RA Puangthong M., Helmann J.D.;  
RT "The OhrR repressor senses organic hydroperoxides by reversible  
RT formation of a cysteine-sulfenic acid derivative.";,  
RC Proc. Natl. Acad. Sci. U.S.A. 99:6690-6695 (2002).  
RA -!- FUNCTION: Organic peroxide sensor. Represses the expression of the  
RT peroxide-inducible gene ohrA by cooperative binding to two  
CC inverted repeat elements.  
CC -!- ENZYME REGULATION: Inactivated by oxidation of Cys-15 to a  
CC sulfenic acid.  
CC -!- SUBUNIT: Homodimer.  
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.  
CC -!- PTM: Oxidation on Cys-15 in response to redox signaling leads to  
CC the loss of DNA-binding activity and the inactivation of repressor  
CC function. Oxidized OhrR can be further reduced by thiol reductants  
CC and repressor activity restored.  
CC -!- SIMILARITY: BELONGS TO THE MARR FAMILY OF TRANSCRIPTIONAL  
CC REGULATORS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).  
CC -----  
DR EMBL; AJ002571; CRA05594.1; -  
DR EMBL; Z99110; CAB13172.1; --  
DR FIR; E69857; E69857.  
DR Subtilist; BG13239; ohrR.  
DR InterPro; IPRO00835; HTH\_MarrR.  
DR Pfam; PF01047; MarrR\_1.  
DR PRINTS; PR00598; HTHMARR.  
DR SMART; SMC0347; HTH\_MARR; 1.  
DR PROSITE; PS01117; HTH\_MARR\_FAMILY; FALSE NEG.  
KW DNA-binding; Transcription regulation; Repressor; Oxidation;  
Complete proteome.  
FT MOD RES 15 15 CYSINE SULFENIC ACID (-SOH) (PROBABLE).  
FT MUTAGEN 15 15 C->G,S: FULL REPRESSOR ACTIVITY, BUT NO  
FT MUTAGEN 15 15 MODULATION BY PEROXIDE.  
SQ SEQUENCE 147 AA; 17003 MW; 4F277EA9AB5EBE861 CRC64;

Query Match 14.1%; Score 81; DB 1; Length 147;  
Best Local Similarity 24.7%; Pred. No. 4.1;  
Matches 20; Conservative 25; Mismatches 34; Indels 2; Gaps 1;  
OV 16 QVKKFRDTRKFNFLNVEIYLNLHLSSESSISKEIAKCFEPYUUKAQKLKDL 75

16 OVKKFFRDTKKKFFNLNYYEEIYILNHILRSESNEISSKEIAKCSEFKPYYLTALQKLDL 75



```

DR PIR; H36845; H36845.
DR PIR; T28521; T28521.
DR InterPro; IPR007066; RNA_pol_A.
DR InterPro; IPR007066; RNA_pol_Rpb1_3.
DR InterPro; IPR007083; RNA_pol_Rpb1_4.
DR InterPro; IPR007083; RNA_pol_Rpb1_4.
DR InterPro; IPR007081; RNA_pol_Rpb1_5.
DR InterPro; IPR006592; RNA_pol_A_N.
DR Pfam; PF00623; RNA_pol_Rpb1_2; 1.
DR Pfam; PF04983; RNA_pol_Rpb1_3; 1.
DR Pfam; PF05000; RNA_pol_Rpb1_4; 1.
DR Pfam; PF04998; RNA_pol_Rpb1_5; 1.
DR SMART; SM00663; RPOA_N; 1.
KW Transferase; DNA-directed RNA polymerase; Transcription;
KW Early protein.
SQ SEQUENCE 1286 AA; 146782 MW; EE9965ACBIDFA93B CRC64;

Query Match      14.1%; Score 81; DB 1; Length 1286;
Best Local Similarity 30.8%; Pred. No. 42;
Matches 33; Conservative 16; Mismatches 20; Indels 38; Gaps 6;

QY 33 EEIYLNHI-----LRSE---SNEISSKEIAKCEPKPYLYTVALOKLKLKLSKRS 83
DB 81 EIRLLNHICHGILRSREPYNDINKELS-----GHARRLKD-KILSKKS 129
QY 84 -----LQDERTVIVVTDTKANI-----QKLISELEY 112
DB 130 CWNSECMQPYQKISFSKKKCFVNKLDIDINVPNSLIYQKLISIEHKP 176

```

Search completed: November 17, 2003, 12:43:37  
 Job time : 18 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.  
OM protein - protein search, using sw model  
Run on: November 17, 2003, 12:24:25 ; Search time 52 Seconds  
(without alignments)  
570.693 Million cell updates/sec  
Title: US-10-043-539A-2  
Perfect score: 574  
Sequence: 1 MSKINDINDLVNATFQVKKF.....DTQXANIQLISELEEVYIKN 115  
Scoring table: BLOSUM62  
Searched: 830525 seqs, 258052604 residues  
Total number of hits satisfying chosen parameters: 830525  
Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries  
Database : SPTREMBL\_23:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_rvirus:\*  
16: sp\_bacteriap:\*  
17: sp\_archaeap:\*  
Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	574	100.0	115	16 Q9F0R1	Q9F0R1 staphylococ
2	486	84.7	114	16 Q8CNC4	Q8CNC4 staphylococ
3	148	25.8	124	16 Q53600	Q53600 staphylococ
4	146	25.4	113	2 Q53777	Q53777 staphylococ
5	143	24.9	124	16 Q85233	Q85233 staphylococ
6	123	21.4	153	16 Q99T44	Q99T44 staphylococ
7	123	21.4	247	2 Q9EZK4	Q9EZK4 staphylococ
8	123	21.4	247	16 Q99RDS	Q99RDS staphylococ
9	120	20.9	99	16 Q8CNU6	Q8CNU6 staphylococ
10	120	20.9	166	2 Q9RFJ6	Q9RFJ6 staphylococ
11	115	20.0	250	16 Q9KMJ2	Q9KMJ2 staphylococ
12	105.5	18.4	1777	5 Q813P4	Q813P4 plasmodium
13	103.5	18.0	131	17 Q96XF0	Q96XF0 sulfolobus
14	101.5	17.7	116	16 Q99S05	Q99S05 staphylococ
15	101	17.6	153	16 Q8PQF1	Q8PQF1 xanthomonas
16	99.5	17.3	242	16 Q8CRE5	Q8CRE5 staphylococ

17	98	17.1	153	2 Q93R11	Q93R11 xanthomonas
18	96	16.7	143	16 Q8XLT0	Q8XLT0 clostridium
19	94.5	16.5	141	16 Q8Y8B9	Q8Y8B9 listeria mo
20	94.5	16.5	147	16 Q8CQ27	Q8CQ27 staphylococ
21	94.5	16.5	154	17 Q978Z3	Q978Z3 thermoplas
22	94	16.4	150	16 Q92CI3	Q92CI3 listeria in
23	93.5	16.3	145	16 Q9CDU5	Q9CDU5 lactococcus
24	93.5	16.3	147	16 Q99VT5	Q99VT5 staphylococ
25	92.5	16.1	148	16 Q8EYB8	Q8EYB8 leptospira
26	92	16.0	247	16 Q99RX6	Q99RX6 staphylococ
27	92	16.0	247	16 Q8NV85	Q8NV85 staphylococ
28	92	16.0	434	5 Q8I386	Q8I386 plasmodium
29	92	16.0	1185	16 Q8XTP0	Q8XTP0 clostridium
30	91.5	15.9	656	5 Q8I4T0	Q8I4T0 plasmodium
31	91	15.9	153	16 Q8FDS2	Q8FDS2 xanthomonas
32	91	15.9	513	5 Q8I326	Q8I326 plasmodium
33	90.5	15.8	145	2 P70734	P70734 acinetobact
34	90.5	15.8	148	2 Q50574	Q50574 bacillus ps
35	90	15.7	157	16 Q92QQ1	Q92QQ1 rhizobium m
36	90	15.7	628	16 Q8D349	Q8D349 wigglewort
37	90	15.7	2166	16 Q51465	Q51465 borelia bu
38	89	15.5	1697	5 Q8IFM4	Q8IFM4 plasmodium
39	89	15.5	1711	5 Q8MWP2	Q8MWP2 plasmodium
40	89	15.5	1713	5 Q8MWP1	Q8MWP1 plasmodium
41	89	15.5	1716	5 Q8MWH2	Q8MWH2 plasmodium
42	88.5	15.4	116	16 Q8CRF1	Q8CRF1 staphylococ
43	88.5	15.4	141	16 Q92D32	Q92D32 listeria in
44	88.5	15.4	921	16 Q8F6A0	Q8F6A0 fusobacteri
45	88.5	15.4	967	2 Q54123	Q54123 staphylococ

ALIGNMENTS

RESULT 1

Q9F0R1	ID	Q9F0R1	PRELIMINARY;	PRT;	115 AA.
AC	Q9F0R1				
DT	01-MAR-2001	(T-EMBLrel. 16, Created)			
DT	01-MAR-2001	(T-EMBLrel. 16, Last sequence update)			
DT	01-OCT-2002	(T-EMBLrel. 22, Last annotation update)			
DE	SRR (Staphylococcal accessory regulator A homolog)				
GN	SRR OR SAV2295 OR SA2089 OR MW2213.				
OS	Staphylococcus aureus (strain Mu50 / ATCC 700699),				
OS	Staphylococcus aureus (strain N315),				
OS	Staphylococcus aureus, and				
OS	Staphylococcus aureus (strain MW2).				
OC	Bacteria; Firmicutes; Bacillales; Staphylococcus.				
OX	NCBI_TaxID=158878, 158879, 1280, 196620;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=RN6390;				
RA	Cheung A.L., Manna A.C.;				
RT	"Characterization of sarR, a modulator of sar expression in				
RT	Staphylococcus aureus";				
RL	Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=Mu50, and N315;				
RX	MEDLINE=21311952; PubMed=11418146;				
RA	Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,				
RA	Cui L., Oguchi A., Aoki K.-I., Nagai Y., Lian J.-Q., Ito T.,				
RA	Kanamori M., Matsumaru H., Maruyama A., Murakami H., Hosoyama A.,				
RA	Mizutani-Ui Y., Takahashi N.K., Sawano T., Inoue R.-I., Kaito C.,				
RA	Sekimizu K., Hirakawa H., Kuhara S., Goto S., Yabuzaki J.,				
RA	Kanehisa M., Yamashita A., Oshima K., Furuya K., Yoshida T.,				
RA	Hattori M., Ogasawara N., Hayashi H., Hiramatsu K.;				
RT	"Whole genome sequencing of methicillin-resistant Staphylococcus				
RT	aureus." Lancet 357:1225-1240 (2001).				
RN	[3]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=MM2;				

```

RX MEDLINE=22040717; PubMed=1204378;
RA Baba T., Takeuchi F., Kuroda M., Yuzawa H., Aoki K.-I., Oguchi A.,
RA Nagai Y., Iwano N., Asano K., Naimi T., Kuroda H., Cui L.,
RA Yamamoto K., Hiramatsu K.;
RT "Genome and virulence determinants of high virulence community-
acquired MRSA.";
RL Lancet 359:1819-1827(2002).
DR EMBL; AF207701; AAG35715.1; -
DR EMBL; AF003364; BAB88457.1; -
DR EMBL; AP003136; BAB43387.1; -
DR EMBL; AP004829; BAB96078.1; -
KW Complete proteome.
SQ SEQUENCE 115 AA; 13669 MW; D2CE40E2DB234DBD CRC64;

Query Match 100.0%; Score 574; DB 16; Length 115;
Best Local Similarity 100.0%; Pred. No. 1.5e-36;
Matches 115; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSKINDINDLVNATQVKKFPRDTKKKNLYEYIILNHLRSNESISKEIAKCSF 60
DB 1 MSKINDINDLVNATQVKKFPRDTKKKNLYEYIILNHLRSNESISKEIAKCSF 60

QY 61 KPYVLTALQKLKDLKLSKRSQDERTVIYVYVDTQKANIQLISELEYIK 115
DB 61 KPYVLTALQKLKDLKLSKRSQDERTVIYVYVDTQKANIQLISELEYIK 115

RESULT 2
Q8CNC4 PRELIMINARY; PRT; 114 AA.
AC Q8CNC4;
DT 01-WAR-2003 (TRENBLrel. 23, Created)
DT 01-WAR-2003 (TRENBLrel. 23, Last sequence update)
DT 01-WAR-2003 (TRENBLrel. 23, Last annotation update)
DE SarR protein.
GN SE1868.
OS Staphylococcus epidermidis.
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=1282;
RN SEQUENCE FROM N.A.
RC STRAIN=ATCC 12228;
RA Zhang Y., Ren S., Li H., Fu G., Lu L., Lu G., Jia J., Tu Y., Qin Z.,
RA Chen Z., Wen Y.;
RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE016750; AAO05509.1; -
KW Complete proteome.
SQ SEQUENCE 114 AA; 13590 MW; A401B6F9F6BECAB CRC64;

Query Match 84.7%; Score 486; DB 16; Length 114;
Best Local Similarity 84.2%; Pred. No. 6.9e-30;
Matches 96; Conservative 8; Mismatches 10; Indels 0; Gaps 0;

QY 1 MSKINDINDLVNATQVKKFPRDTKKKNLYEYIILNHLRSNESISKEIAKCSF 60
DB 1 MSKINDINDLVNATQVKKFPRDTKKKNLYEYIILNHLRSNESISKEIAKCSF 60

QY 61 KPYVLTALQKLKDLKLSKRSQDERTVIYVYVDTQKANIQLISELEYIK 114
DB 61 KPYVLTALQKLKDLNLLSKRSVHDETVIVFVSDQEKIKLILELENYIK 114

RESULT 3
Q53600 PRELIMINARY; PRT; 124 AA.
AC Q53600;
DT 01-NOV-1996 (TRENBLrel. 01, Created)
DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)
DT 01-OCT-2002 (TRENBLrel. 22, Last annotation update)
DE Staphylococcal accessory regulator variant (Staphylococcal
DE regulator A).
GN SARA OR SAV0616 OR SA0573 OR MW0580.
OS Staphylococcus aureus,

```

```

OS Staphylococcus aureus (strain Mu50 / ATCC 700699),
OS Staphylococcus aureus (strain N315), and
OS Staphylococcus aureus (strain MW2).
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=1280, 158878, 158879, 196820;
RN SEQUENCE FROM N.A.
RC SPECIES=S.aureus; STRAIN=RM450;
RA Shawcross S.G., Edwards-Jones V., Dawson M.M., Foster H.A.;
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
RN SEQUENCE FROM N.A.
RC SPECIES=S.aureus; STRAIN=RM450;
RX MEDLINE=94292439; PubMed=80211198;
RA Cheung A.L., Projan S.J.;
RT "Cloning and sequencing of sarA of Staphylococcus aureus, a gene
required for the expression of agr.";
RL J. Bacteriol. 176:4168-4172(1994).
RN SEQUENCE FROM N.A.
RC SPECIES=S.aureus; STRAIN=RM450;
RX MEDLINE=92335318; PubMed=1321441;
RA Cheung A.L., Koomey J.M., Butler C.A., Projan S.J., Fischetti V.A.;
RT "Regulation of exoprotein expression in Staphylococcus aureus by a
locus (sar) distinct from agr.";
RL Proc. Natl. Acad. Sci. U.S.A. 89:6462-6466(1992).
RN SEQUENCE FROM N.A.
RC SPECIES=S.aureus; STRAIN=RM450;
RA Cheung A.L.;
RL Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.
RN SEQUENCE FROM N.A.
RC SPECIES=S.aureus; STRAIN=Mu50, and N315;
RX MEDLINE=21311952; PubMed=11418146;
RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,
RA Cui L., Oguchi A., Aoki K.-I., Nagai Y., Lian J.-Q., Ito T.,
RA Kanamori M., Matsumaru H., Maruyama A., Murakami H., Hoshoyama A.,
RA Mizutani-Ui Y., Takahashi N.K., Sawano T., Inoue R.-I., Kaito C.,
RA Sekimizu K., Hirakawa H., Kihara S., Goto S., Yabuzaki J.,
RA Kanehisa M., Ogasawara N., Hayashi H., Hiramatsu K.,
RA Hattori M., Ogasawara N., Hayashi H., Hiramatsu K.;
RT "Whole genome sequencing of methicillin-resistant Staphylococcus
aureus.";
RL Lancet 357:1225-1240(2001).
RN SEQUENCE FROM N.A.
RC SPECIES=S.aureus; STRAIN=MM2;
RX MEDLINE=22040717; PubMed=1204378;
RA Baba T., Takeuchi F., Kuroda M., Yuzawa H., Aoki K.-I., Oguchi A.,
RA Nagai Y., Iwano N., Asano K., Naimi T., Kuroda H., Cui L.,
RA Yamamoto K., Hiramatsu K.;
RT "Genome and virulence determinants of high virulence community-
acquired MRSA.";
RL Lancet 359:1819-1827(2002).
DR EMBL; AF515775; AAM74164.1; -
DR EMBL; U20782; AAA62477.1; -
DR EMBL; AP003359; BAB56778.1; -
DR EMBL; AP003131; BAB41805.1; -
DR EMBL; AP004824; BAB94445.1; -
KW Complete proteome.
SQ SEQUENCE 124 AA; 14718 MW; DB9A16E806C10661 CRC64;

Query Match 25.8%; Score 148; DB 16; Length 124;
Best Local Similarity 30.1%; Pred. No. 0.00032;
Matches 34; Conservative 27; Mismatches 52; Indels 0; Gaps 0;

QY 1 MSKINDINDLVNATQVKKFPRDTKKKNLYEYIILNHLRSNESISKEIAKCSF 60
DB 3 ITKINDCFELSMVYADKLKSLIKKFSISFEFAVLITYISENKEKYKDIINHLNY 62
QY 61 KPYVLTALQKLKDLKLSKRSQDERTVIYVYVDTQKANIQLISELEYIK 113

```



Db 63 KQPQVVKAVKILSQBDFYDFKRNHEDERTVLILVNAQQKKIESLLSRVVKRI 115

## RESULT 4

Q53777 PRELIMINARY; PRT; 113 AA.  
AC Q53777; (TREMELrel. 01, Created)  
DT 01-NOV-1996 (TREMELrel. 01, Last sequence update)  
DT 01-NOV-1996 (TREMELrel. 01, Last sequence update)  
DT 01-DEC-2001 (TREMELrel. 19, Last annotation update)  
DE SARA.  
GN SARA.  
OS Staphylococcus aureus.  
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.  
OX NCBI\_TaxID=1280;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=RN6390;  
RX MEDLINE=94292439; PubMed=80211198;  
RA Cheung A.L., Projan S.J.;  
RT "The molecular architecture of the sar locus in Staphylococcus aureus";  
RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=RN6390;  
RX MEDLINE=94292439; PubMed=80211198;  
RA Cheung A.L., Projan S.J.;  
RT "Cloning and sequencing of sarA of Staphylococcus aureus, a gene required for the expression of agr.";  
RL J. Bacteriol. 176:4168-4172(1994).  
DR EMBL; U46541; AAB05396.1; -;  
SQ SEQUENCE 113 AA; 13469 MW; 6A6D53F34E010AFB CRC64;

Query Match 25.4%; Score 146; DB 2; Length 113;

Best Local Similarity 29.7%; Pred. No. 0.00041;

Matches 33; Conservative 27; Mismatches 51; Indels 0; Gaps 0;

Qy 1 MSKINDINDLVNATFQVKKFFRDTKKKFNLVVEEYVILNHLRSESNEISKEIAKCSF 60

Db 3 ITKINDCEFLSMVTYADRLKSLIKESISFEFAVLTYISENKEEYVFKDIINHLNY 62

Qy 61 KPYLTKALQKLDLKLKRSLODERTVIVVYDTOKANIQKLISELEE 111

Db 63 KQPQVVKAVKILSQBDFYDFKRNHEDERTVLILVNAQQKKIESLLSRVVK 113

## RESULT 5

O85233 PRELIMINARY; PRT; 124 AA.  
AC O85233;  
DT 01-NOV-1998 (TREMELrel. 08, Created)  
DT 01-NOV-1998 (TREMELrel. 08, Last sequence update)  
DT 01-MAR-2003 (TREMELrel. 23, Last annotation update)  
DE Staphylococcal accessory regulator A homolog.  
GN SARA OR SE0390.  
OS Staphylococcus epidermidis.  
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.  
OX NCBI\_TaxID=1282;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=6937;  
RA Fluckiger U., Wolz C., Cheung A.L.;  
RT "Characterization of a sar homolog of Staphylococcus epidermidis";  
RL Infect. Immun. 0:0-0(1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 12228;  
RA Zhang Y., Ren S., Li H., Fu G., Lu L., Lu G., Jia J., Tu Y., Qin Z.,  
RA Chen Z., Wen Y.;  
RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF054173; AAC25106.1; -;  
DR EMBL; AE016745; AAO03987.1; -;  
KW Complete proteome.

SQ SEQUENCE 124 AA; 14731 MW; 4D1D10E47D574266 CRC64;

Query Match 24.9%; Score 143; DB 16; Length 124;

Best Local Similarity 29.2%; Pred. No. 0.00076;

Matches 33; Conservative 25; Mismatches 55; Indels 0; Gaps 0;

Qy 1 MSKINDINDLVNATFQVKKFFRDTKKKFNLVVEEYVILNHLRSESNEISKEIAKCSF 60

Db 3 ISKINDCFELLAMVTYADRLKGLIKESISFEFAVLTYISENKEEYVFKDIINHLNY 62

Qy 61 KPYLTKALQKLDLKLKRSLODERTVIVVYDTOKANIQKLISEEEYI 113

Db 63 KQPQVVKAVKILSQBDFYDFKRNHEDERTVLILVDSKQRKDDLLKRVNRI 115

## RESULT 6

Q99TA4 PRELIMINARY; PRT; 153 AA.  
AC Q99TA4;  
DT 01-JUN-2001 (TREMELrel. 17, Created)  
DT 01-JUN-2001 (TREMELrel. 17, Last sequence update)  
DT 01-OCT-2002 (TREMELrel. 22, Last annotation update)  
DE Repressor of toxins Rot.  
GN ROT OR SAV1764 OR SA1583 OR MW1705.  
OS Staphylococcus aureus (strain Mu50 / ATCC 700699),  
OS Staphylococcus aureus (strain N315), and  
OS Staphylococcus aureus (strain MW2).  
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.  
OX NCBI\_TaxID=158878, 158879, 158879, 158879;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Mu50, and N315;  
RX MEDLINE=21311952; PubMed=11418146;  
RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,  
RA Cui L., Oguchi A., Aoki K.-I., Nagai Y., Lian J.-Q., Ito T.,  
RA Kanamori M., Matsumaru H., Maruyama A., Murakami H., Hosoyama A.,  
RA Mizutani-Ji Y., Takahashi N.K., Sawano T., Inoue R.-I., Kaito C.,  
RA Sekinizu K., Hirakawa H., Kuhara S., Goto S., Yabuzaki J.,  
RA Kanehisa M., Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T.,  
RA Hattori M., Ogasawara N., Hayashi H., Hiramatsu K.;  
RT "Whole genome sequencing of methicillin-resistant Staphylococcus aureus";  
RL Lancet 357:1225-1240(2001).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=MW2;  
RX MEDLINE=22040717; PubMed=12044378;  
RA Baba T., Takeuchi F., Kuroda M., Yuzawa H., Aoki K.-I., Oguchi A.,  
RA Nagai Y., Iwama N., Asano K., Naimi T., Kuroda H., Cui L.,  
RA Yamamoto K., Hiramatsu K.;  
RT "Genome and virulence determinants of high virulence community-acquired MRSA";  
RL Lancet 359:1819-1827(2002).  
DR EMBL; AP003363; BAB57926.1; -;  
DR EMBL; AP003135; BAB42851.1; -;  
DR EMBL; AP004828; BAB95570.1; -;  
KW Complete proteome.  
SQ SEQUENCE 153 AA; 18009 MW; E24C6E9BCC353E20 CRC64;

Query Match 21.4%; Score 123; DB 16; Length 153;

Best Local Similarity 28.0%; Pred. No. 0.03;

Matches 33; Conservative 26; Mismatches 45; Indels 14; Gaps 4;

Qy 1 MSKINDINDLVNATFQVKKFFRDTKKKFNLVVEEYVILNHLRSESNEISKEI 53

Db 21 MKKVN-NDTVFGILQETLLGDINSIFSEISEYKMSREELIL-LTLWQKGMTLXE 76

Qy 54 IAKCSFKPYLTALQKLDLKLKRSLODERTVIVVYDTOKANIQKLISE 108

Db 77 MDRFVEVKPYKRTYTNLNVLEWLEVIYKRPVDDERTVIIHNEKLOQEKVELLNFI 134

## RESULT 7



```
RL J. Bacteriol. 182:3197-3203(2000).
DR EMBL; AF182239; AAF22306.1; -- 251AE45481E5699D CRC64;
SQ SEQUENCE 166 AA; 19431 MW; 251AE45481E5699D CRC64;

Query Match
Best Local Similarity 20.9%; Score 120; DB 2; Length 166;
Matches 33; Conservative 26; Mismatches 45; Indels 14; Gaps 4;

QY 1 MSKINDINDLVNATQ-----VKFPRDTKKENLVYIYLHNLHLSSESNEISKE 53
Db 34 MKKVN--NDTVFGILQLETLGDINSIFSEISEYKMSREEIIL--ITLQKGFMTLKE 89

QY 54 IACSEFPYVLTALQKLDKLLSKRSQDERTVIVYVD---TOKANIQLISE 108
Db 90 MDRFEVVPKXETRYNNLVLELWYKRPVYDERTVVIHFNKLOQKVELLNFISD 147

RESULT 11
Q9KJW2 PRELIMINARY; PRT; 250 AA.
AC Q9KJW2;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Hypothetical protein (SaxH1 protein) (Staphylococcal accessory
DE regulator A homologue).
GN CRFX OR SAXH1 OR SAV0112 OR SA0108 OR MW0085.
OS Staphylococcus aureus (strain Mu50 / ATCC 700699),
OS Staphylococcus aureus (strain N315),
OS Staphylococcus aureus, and
OS Staphylococcus aureus (strain MW2).
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=158878, 158879, 1280, 196620;
RN [1]_
RP SEQUENCE FROM N.A.
RC STRAIN=MU3;
RX MEDLINE=20175229; PubMed=10709580;
RA Kuroda M., Kuwahara-Arai K., Hiramatsu K.;
RT "Identification of the up- and down-regulated genes in vancomycin-
RT resistant Staphylococcus aureus strains Mu3 and Mu50 by cDNA
RT differential hybridization method.";
RL Biochem. Biophys. Res. Commun. 265:485-490(2000).
RN [2]_
RP SEQUENCE FROM N.A.
RC STRAIN=N315, and Mu50 / ATCC 700699;
RX MEDLINE=21311952; PubMed=11418146;
RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,
RA Cui L., Oguchi A., Aoki K.-I., Nagai Y., Lian J.-Q., Ito T.,
RA Kanamori M., Matsumaru H., Maruyama A., Murakami H., Hosoyana A.,
RA Mizutani-Uji Y., Takanashi N.K., Sawano T., Inoue R.-I., Kaito C.,
RA Sekimizu K., Hirakawa H., Kuhara S., Goto S., Yabuzaki J.,
RA Kanahisa M., Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T.,
RA Hattori M., Ogasawara N., Hayashi H., Hiramatsu K.;
RT "Whole genome sequencing of methicillin-resistant Staphylococcus
RT aureus.";
RL Lancet 357:1225-1240(2001).
RN [3]_
RP SEQUENCE FROM N.A.
RC STRAIN=N315;
RA Kuroda-Murakami H., Kuroda M., Hiramatsu K.;
RT "Identification of differentially expressed genes of Staphylococcus
RT aureus in response to and in raised resistance to imipenem.";
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
RN [4]_
RP SEQUENCE FROM N.A.
RC STRAIN=MW2;
RX MEDLINE=22040717; PubMed=12044378;
RA Baba T., Takeuchi F., Kuroda M., Yuzawa H., Aoki K.-I., Oguchi A.,
RA Nagai Y., Iwana N., Asano K., Naini T., Kuroda H., Cui L.,
RA Yamamoto K., Hiramatsu K.;
RT "Genome and virulence determinants of high virulence community-
RT acquired MRSA.";
RL Lancet 359:1819-1827(2002).

DR EMBL; AB035454; BAB03341.1; --
DR EMBL; AP003129; BAB1327.1; --
DR EMBL; AB050858; BAB9826.1; --
DR EMBL; AP003358; BAB56274.1; --
DR EMBL; AP004822; BAB93950.1; --
KW Complete proteome.
SQ SEQUENCE 250 AA; 29890 MW; 4680D50FE86DDC19 CRC64;

Query Match
Best Local Similarity 20.0%; Score 115; DB 16; Length 250;
Matches 27; Conservative 26; Mismatches 55; Indels 0; Gaps 0;

QY 6 DINDLVNATFQVKFPRDTKKENLVYIYLHNLHLSSESNEISKEIACSEFPYVL 65
Db 132 DSKEFLNMYTFYFNKILKHLTSLFVEFTLAIITTSQNKVILKDLIETHHKYPQT 191

QY 66 TKALQKLDKLLSKRSQDERTVIVYVD---TOKANIQLISELEYI 113
Db 192 VRALNNLKQGYLIKERTSTEDERKILHMDAQODHAEQLLAQVNL 239

RESULT 12
Q813P4 PRELIMINARY; PRT; 1777 AA.
AC Q813P4;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical protein.
GN PFE1095W.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]_
RP SEQUENCE FROM N.A.
RA Devlin K., Baker S., Davies P., Mungal K., Berriman M., Pain A.,
RA Hall N., Bowman S., Churcher C., Quail M., Barrell B.;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RN [2]_
RP SEQUENCE FROM N.A.
RX MEDLINE=22255709; PubMed=12368867;
RA Hall N., Pain A., Berriman M., Churcher C., Harris B., Harris D.,
RA Mungall K., Bowman S., Atkin R., Baker S., Barron A., Brooks K.,
RA Buckee C.O., Burrows C., Cherevach I., Chillingworth C.,
RA Chillingworth T., Christodoulou Z., Clark L., Clark R., Corton C.,
RA Cronin A., Davies R., Davis P., Dear P., Dearden F., Doggett J.,
RA Feltwell T., Goble A., Goodhead I., Gwilliam R., Hamlin N., Hance Z.,
RA Harper D., Hauser H., Hornsby T., Holtroyd S., Horrocks P.,
RA Humphray S., Jagels K., James K.D., Johnson D., Kerhornou A.,
RA Knights A., Konfortov B., Kyes S., Larke N., Lawson D., Lennard N.,
RA Line A., Maddison M., Mclean J., Mooney P., Moule S., Murphy L.,
RA Oliver K., Ormond D., Price C., Quail M.A., Rabinowitsch E.,
RA Rajandream M.A., Rutter S., Rutherford K.M., Sanders M., Simmonds M.,
RA Seeger K., Sharp S., Smith R., Squares R., Squares S., Stevens K.,
RA Taylor K., Tivey A., Unwin L., Whitehead S., Woodward J.,
RA Sulston J.E., Craig A., Newbold C., Barrell B.G.;
RT "Sequence of Plasmodium falciparum chromosomes 1, 3-9 and 13.";
RL Nature 419:527-531(2002).
DR EMBL; AL929353; CAB51583.1; --
KW Hypothetical protein.
SQ SEQUENCE 1777 AA; 213320 MW; 244467CFF190C522 CRC64;

Query Match
Best Local Similarity 18.4%; Score 105.5; DB 5; Length 1777;
Matches 33; Conservative 24; Mismatches 32; Indels 29; Gaps 5;

QY 5 NDINDLVNATFQVKFPRDTKKENLVYIYLHNLHLSSESNEISKEIAC 57
Db 1289 NDINEKVLKQREEIFYEKKNFEKNEHLKLVNKNVIKIDRENLNKE--- 1344

QY 58 SEKPYVLTALQKLDKLLSKRSQDERTVIVYVD---TOKANIQLISELEYI 114
Db 1345 -----BELKMKKEI-LLSKE-----NIINYSLSNKKINKKLEYNKLEDTK 1385
```



Search completed: November 17, 2003, 12:44:37  
Job time : 56 secs

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 17, 2003, 12:43:21 ; Search time 159 Seconds  
(without alignments)  
658.117 Million cell updates/sec

Title: US-10-043-539A-2

Perfect score: 574

Sequence: 1 MSKINDINDLVNATFQVKF.....DTQXNIQKLISELEEVYKN 115

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 5728757 seqs, 909918778 residues

Total number of hits satisfying chosen parameters: 5728757

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Pending Patents AA Main:\*

- 1: /cgn2\_6/ptodata/2/paa/PCTUS COMB.pcp.\*
- 2: /cgn2\_6/ptodata/2/paa/US06 COMB.pcp.\*
- 3: /cgn2\_6/ptodata/2/paa/US07 COMB.pcp.\*
- 4: /cgn2\_6/ptodata/2/paa/US08 COMB.pcp.\*
- 5: /cgn2\_6/ptodata/2/paa/US081 COMB.pcp.\*
- 6: /cgn2\_6/ptodata/2/paa/US082 COMB.pcp.\*
- 7: /cgn2\_6/ptodata/2/paa/US083 COMB.pcp.\*
- 8: /cgn2\_6/ptodata/2/paa/US084 COMB.pcp.\*
- 9: /cgn2\_6/ptodata/2/paa/US085 COMB.pcp.\*
- 10: /cgn2\_6/ptodata/2/paa/US086 COMB.pcp.\*
- 11: /cgn2\_6/ptodata/2/paa/US087 COMB.pcp.\*
- 12: /cgn2\_6/ptodata/2/paa/US088 COMB.pcp.\*
- 13: /cgn2\_6/ptodata/2/paa/US089 COMB.pcp.\*
- 14: /cgn2\_6/ptodata/2/paa/US090 COMB.pcp.\*
- 15: /cgn2\_6/ptodata/2/paa/US091 COMB.pcp.\*
- 16: /cgn2\_6/ptodata/2/paa/US092 COMB.pcp.\*
- 17: /cgn2\_6/ptodata/2/paa/US093 COMB.pcp.\*
- 18: /cgn2\_6/ptodata/2/paa/US094 COMB.pcp.\*
- 19: /cgn2\_6/ptodata/2/paa/US095 COMB.pcp.\*
- 20: /cgn2\_6/ptodata/2/paa/US096 COMB.pcp.\*
- 21: /cgn2\_6/ptodata/2/paa/US097 COMB.pcp.\*
- 22: /cgn2\_6/ptodata/2/paa/US098 COMB.pcp.\*
- 23: /cgn2\_6/ptodata/2/paa/US099 COMB.pcp.\*
- 24: /cgn2\_6/ptodata/2/paa/US099A COMB.pcp.\*
- 25: /cgn2\_6/ptodata/2/paa/US099B COMB.pcp.\*
- 26: /cgn2\_6/ptodata/2/paa/US100 COMB.pcp.\*
- 27: /cgn2\_6/ptodata/2/paa/US101 COMB.pcp.\*
- 28: /cgn2\_6/ptodata/2/paa/US102 COMB.pcp.\*
- 29: /cgn2\_6/ptodata/2/paa/US103 COMB.pcp.\*
- 30: /cgn2\_6/ptodata/2/paa/US104 COMB.pcp.\*
- 31: /cgn2\_6/ptodata/2/paa/US106 COMB.pcp.\*
- 32: /cgn2\_6/ptodata/2/paa/US60 COMB.pcp.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	574	100.0	115	1	PCT-US02-00877-2 Sequence 2, Appli

2	574	100.0	115	26	US-10-043-539-2	Sequence 2, Appli
3	574	100.0	115	26	US-10-043-539A-2	Sequence 2, Appli
4	574	100.0	129	12	US-08-827-356-4882	Sequence 4882, Ap
5	574	100.0	129	12	US-09-611-529-6523	Sequence 6523, Ap
6	574	100.0	129	25	US-09-950-084-6523	Sequence 6523, Ap
7	486	84.7	128	18	US-09-450-969-6487	Sequence 6487, Ap
8	486	84.7	128	26	US-10-092-411A-4993	Sequence 4993, Ap
9	224	39.0	120	1	PCT-US02-00877-26	Sequence 26, Appl
10	224	39.0	120	26	US-10-043-539A-26	Sequence 26, Appl
11	224	39.0	120	26	US-10-043-539A-26	Sequence 26, Appl
12	148	25.8	120	1	PCT-US02-00877-35	Sequence 35, Appl
13	148	25.8	120	26	US-10-043-539A-35	Sequence 35, Appl
14	148	25.8	124	1	PCT-US02-00877-3	Sequence 3, Appli
15	148	25.8	124	20	US-09-612-549A-1	Sequence 1, Appli
16	148	25.8	124	26	US-10-043-539-3	Sequence 3, Appli
17	148	25.8	124	26	US-10-043-539A-3	Sequence 3, Appli
18	148	25.8	124	28	US-10-290-142-9	Sequence 9, Appli
19	148	25.8	124	28	US-10-290-142-8	Sequence 8, Appli
20	148	25.8	147	20	US-09-611-529-5183	Sequence 5183, Ap
21	148	25.8	147	25	US-09-950-084-5183	Sequence 5183, Ap
22	145	25.3	149	26	US-10-092-411A-5151	Sequence 5151, Ap
23	143	24.9	124	20	US-09-612-549A-2	Sequence 2, Appli
24	143	24.9	148	18	US-09-450-969-7381	Sequence 7381, Ap
25	143	24.9	148	26	US-10-092-411A-5614	Sequence 5614, Ap
26	138	24.0	149	18	US-09-450-969-6116	Sequence 6116, Ap
27	130	22.6	124	20	US-09-612-549A-19	Sequence 19, Appl
28	123	21.4	247	27	US-10-145-602-4	Sequence 4, Appli
29	123	21.4	261	12	US-08-827-356-5491	Sequence 5491, Ap
30	123	21.4	261	20	US-09-611-529-7218	Sequence 7218, Ap
31	123	21.4	261	25	US-09-950-084-7218	Sequence 7218, Ap
32	120	20.9	166	27	US-10-145-602-2	Sequence 2, Appli
33	115	20.0	118	1	PCT-US02-00877-28	Sequence 28, Appl
34	115	20.0	118	26	US-10-043-539-28	Sequence 28, Appl
35	115	20.0	118	26	US-10-043-539A-28	Sequence 28, Appl
36	115	20.0	260	12	US-08-827-356-4903	Sequence 4903, Ap
37	115	20.0	260	20	US-09-611-529-6537	Sequence 6537, Ap
38	115	20.0	260	25	US-09-950-084-6537	Sequence 6537, Ap
39	113	19.7	121	12	US-08-827-356-4158	Sequence 4158, Ap
40	113	19.7	121	20	US-09-611-529-5857	Sequence 5857, Ap
41	113	19.7	121	25	US-09-950-084-5857	Sequence 5857, Ap
42	101.5	17.7	131	12	US-08-827-356-3856	Sequence 3856, Ap
43	101.5	17.7	131	20	US-09-611-529-6519	Sequence 6519, Ap
44	101.5	17.7	131	25	US-09-950-084-6519	Sequence 6519, Ap
45	101.5	17.7	162	16	US-09-252-691-7968	Sequence 7968, Ap

ALIGNMENTS

RESULT 1

PCT-US02-00877-2  
; Sequence 2, Application PC/TUS0200877  
; GENERAL INFORMATION:  
; APPLICANT: Trustees of Dartmouth College  
; TITLE OF INVENTION: Compositions and Methods for affecting virulence determinants in  
; TITLE OF INVENTION: bacteria  
; FILE REFERENCE: 11312-006-228  
; CURRENT APPLICATION NUMBER: PCT/US02/00877  
; PRIOR FILING DATE: 2003-03-04  
; PRIOR APPLICATION NUMBER: 60/261,233  
; PRIOR FILING DATE: 2001-01-12  
; PRIOR APPLICATION NUMBER: 60/261,607  
; PRIOR FILING DATE: 2001-01-12  
; PRIOR APPLICATION NUMBER: 60/389,601  
; PRIOR FILING DATE: 2001-05-08  
; NUMBER OF SEQ ID NOS: 35  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 2  
; LENGTH: 115  
; TYPE: PRT  
; ORGANISM: Staphylococcus aureus  
PCT-US02-00877-2

```
Query Match      100.0%; Score 574; DB 1; Length 115;
Best Local Similarity 100.0%; Pred. No. 3.6e-49;
Matches 115; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSKINDINDLVNATQVKKFFRDTKKKFNLYEEIYILNHLRSSEISKEIAKCSF 60
DB 1 MSKINDINDLVNATQVKKFFRDTKKKFNLYEEIYILNHLRSSEISKEIAKCSF 60

QY 61 KPYLTALQKLKDLKLSKRSLODERTVIYVYVDTQKANIQLISELEYIKN 115
DB 61 KPYLTALQKLKDLKLSKRSLODERTVIYVYVDTQKANIQLISELEYIKN 115

RESULT 2
US-10-043-539-2
; Sequence 2, Application US/10043539
; GENERAL INFORMATION:
; APPLICANT: Cheung, Ambrose
; APPLICANT: Manna, Adar
; APPLICANT: Zhang, Gorgyi
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR AFFECTING VIRULENCE DETERMINANTS IN
; FILE REFERENCE: DC-0199
; CURRENT APPLICATION NUMBER: US/10/043.539
; PRIOR FILING DATE: 2002-01-11
; PRIOR APPLICATION NUMBER: US 60/261.233
; PRIOR FILING DATE: 2001-01-12
; PRIOR APPLICATION NUMBER: US 60/261.607
; PRIOR FILING DATE: 2001-01-12
; PRIOR APPLICATION NUMBER: US 60/289.601
; PRIOR FILING DATE: 2001-05-08
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 115
; TYPE: PRT
; ORGANISM: Staphylococcus aureus -
US-10-043-539-2

Query Match      100.0%; Score 574; DB 26; Length 115;
Best Local Similarity 100.0%; Pred. No. 3.6e-49;
Matches 115; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSKINDINDLVNATQVKKFFRDTKKKFNLYEEIYILNHLRSSEISKEIAKCSF 60
DB 1 MSKINDINDLVNATQVKKFFRDTKKKFNLYEEIYILNHLRSSEISKEIAKCSF 60

QY 61 KPYLTALQKLKDLKLSKRSLODERTVIYVYVDTQKANIQLISELEYIKN 115
DB 61 KPYLTALQKLKDLKLSKRSLODERTVIYVYVDTQKANIQLISELEYIKN 115

RESULT 3
US-10-043-539A-2
; Sequence 2, Application US/10043539A
; GENERAL INFORMATION:
; APPLICANT: Cheung, Ambrose
; APPLICANT: Manna, Adar
; APPLICANT: Zhang, Gorgyi
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR AFFECTING VIRULENCE DETERMINANTS IN
; FILE REFERENCE: DC-0199
; CURRENT APPLICATION NUMBER: US/10/043.539A
; CURRENT FILING DATE: 2000-01-11
; PRIOR APPLICATION NUMBER: US 60/261.233
; PRIOR FILING DATE: 2001-01-12
; PRIOR APPLICATION NUMBER: US 60/261.607
; PRIOR FILING DATE: 2001-01-12
; PRIOR APPLICATION NUMBER: US 60/289.601
; PRIOR FILING DATE: 2001-05-08
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 115
```

```
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-10-043-539A-2

Query Match      100.0%; Score 574; DB 26; Length 115;
Best Local Similarity 100.0%; Pred. No. 3.6e-49;
Matches 115; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSKINDINDLVNATQVKKFFRDTKKKFNLYEEIYILNHLRSSEISKEIAKCSF 60
DB 1 MSKINDINDLVNATQVKKFFRDTKKKFNLYEEIYILNHLRSSEISKEIAKCSF 60

QY 61 KPYLTALQKLKDLKLSKRSLODERTVIYVYVDTQKANIQLISELEYIKN 115
DB 61 KPYLTALQKLKDLKLSKRSLODERTVIYVYVDTQKANIQLISELEYIKN 115

RESULT 4
US-08-827-356-4882
; Sequence 4882, Application US/08827356
; GENERAL INFORMATION:
; APPLICANT: George H. Shimer, Jr.
; APPLICANT: George H. Miller
; APPLICANT: Roberta S. Hare
; APPLICANT: Karen J. Shaw
; TITLE OF INVENTION: STAPHYLOCOCCUS AUREUS RELATED
; TITLE OF INVENTION: COMPOSITIONS AND METHODS
; NUMBER OF SEQUENCES: 5574
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Schering-plough Corporation
; STREET: 2000 Galloping Hill Road
; CITY: Kenilworth
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07033-0530
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/827,356
; FILING DATE: 01-APR-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/014,477
; FILING DATE: 01-APR-1996
; APPLICATION NUMBER: 60/016,743
; FILING DATE: 02-MAY-1996
; APPLICATION NUMBER: 60/020,016
; FILING DATE: 14-JUN-1996
; INFORMATION FOR SEQ ID NO: 4882:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 129 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: Protein
; HYPOTHEetical: YES
; ORIGINAL SOURCE:
; ORGANISM: Staphylococcus aureus
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 1...129
US-08-827-356-4882

Query Match      100.0%; Score 574; DB 12; Length 129;
Best Local Similarity 100.0%; Pred. No. 4.2e-49;
Matches 115; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSKINDINDLVNATQVKKFFRDTKKKFNLYEEIYILNHLRSSEISKEIAKCSF 60
DB 15 MSKINDINDLVNATQVKKFFRDTKKKFNLYEEIYILNHLRSSEISKEIAKCSF 74

QY 61 KPYLTALQKLKDLKLSKRSLODERTVIYVYVDTQKANIQLISELEYIKN 115
```



Db 75 KPYLTAKLQKLDKLLSKRSLODERTVIVVYDTQKANTQKLISELEYIKN 129  
|||||

RESULT 5  
US-09-611-529-6523  
; Sequence 6523, Application US/09611529  
; GENERAL INFORMATION:  
; APPLICANT: George H. Shimer, Jr.  
; APPLICANT: George H. Miller  
; APPLICANT: Roberta S. Hare  
; APPLICANT: Karen J. Shaw  
; TITLE OF INVENTION: Staphylococcus aureus Related Compositions and Methods  
; FILE REFERENCE: 1034/IC963U81  
; CURRENT APPLICATION NUMBER: US/09/611,529  
; CURRENT FILING DATE: 2000-06-30  
; PRIOR APPLICATION NUMBER: US 09/417,811  
; PRIOR FILING DATE: 1999-10-14  
; PRIOR APPLICATION NUMBER: US 09/353,718  
; PRIOR FILING DATE: 1999-07-14  
; PRIOR APPLICATION NUMBER: US 09/266,557  
; PRIOR FILING DATE: 1999-03-11  
; PRIOR APPLICATION NUMBER: US 09/266,556  
; PRIOR FILING DATE: 1999-03-11  
; PRIOR APPLICATION NUMBER: US 09/266,555  
; PRIOR FILING DATE: 1999-03-11  
; PRIOR APPLICATION NUMBER: US 09/266,542  
; PRIOR FILING DATE: 1999-03-11  
; PRIOR APPLICATION NUMBER: US 09/266,541  
; PRIOR FILING DATE: 1999-03-11  
; PRIOR APPLICATION NUMBER: US 09/037,934  
; PRIOR FILING DATE: 1998-03-10  
; PRIOR APPLICATION NUMBER: US 09/036,720  
; PRIOR FILING DATE: 1998-03-06  
; PRIOR APPLICATION NUMBER: US 09/036,338  
; PRIOR FILING DATE: 1998-03-06  
; PRIOR APPLICATION NUMBER: US 09/036,334  
; PRIOR FILING DATE: 1998-03-06  
; PRIOR APPLICATION NUMBER: US 09/036,221  
; PRIOR FILING DATE: 1998-03-06  
; PRIOR APPLICATION NUMBER: US 09/036,137  
; PRIOR FILING DATE: 1998-03-06  
; PRIOR APPLICATION NUMBER: US 09/036,082  
; PRIOR FILING DATE: 1998-03-06  
; PRIOR APPLICATION NUMBER: US 09/036,081  
; PRIOR FILING DATE: 1998-03-06  
; PRIOR APPLICATION NUMBER: US 09/036,079  
; PRIOR FILING DATE: 1998-03-06  
; PRIOR APPLICATION NUMBER: US 09/035,913  
; PRIOR FILING DATE: 1998-03-06  
; PRIOR APPLICATION NUMBER: US 09/035,744  
; PRIOR FILING DATE: 1998-03-06  
; PRIOR APPLICATION NUMBER: US 08/827,356  
; PRIOR FILING DATE: 1997-04-01  
; PRIOR APPLICATION NUMBER: US 08/831,156  
; PRIOR FILING DATE: 1997-04-01  
; PRIOR APPLICATION NUMBER: US 60/014,477  
; PRIOR FILING DATE: 1996-04-01  
; PRIOR APPLICATION NUMBER: US 60/016,743  
; PRIOR FILING DATE: 1996-05-02  
; PRIOR APPLICATION NUMBER: US 60/020,016  
; PRIOR FILING DATE: 1996-06-14  
; NUMBER OF SEQ ID NOS: 7451  
; SEQ ID NO 6523  
; LENGTH: 129  
; TYPE: PRT  
; ORGANISM: Staphylococcus aureus  
US-09-611-529-6523

Query Match 100.0%; Score 574; DB 20; Length 129;  
Best Local Similarity 100.0%; Pred. No. 4.2e-49;  
Matches 115; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MSKINDINDLVNATFQVKFFRDTKKFNLYNEEIVILNHLRSENEISSKEIAKCSF 60  
Db 15 MSKINDINDLVNATFQVKFFRDTKKFNLYNEEIVILNHLRSENEISSKEIAKCSF 74  
Qy 61 KPYLTAKLQKLDKLLSKRSLODERTVIVVYDTQKANTQKLISELEYIKN 115  
Db 75 KPYLTAKLQKLDKLLSKRSLODERTVIVVYDTQKANTQKLISELEYIKN 129  
|||||

RESULT 6  
US-09-950-084-6523  
; Sequence 6523, Application US/09950084  
; GENERAL INFORMATION:  
; APPLICANT: George H. Shimer, Jr.  
; APPLICANT: George H. Miller  
; APPLICANT: Roberta S. Hare  
; APPLICANT: Karen J. Shaw  
; TITLE OF INVENTION: Staphylococcus aureus Related Compositions and Methods  
; FILE REFERENCE: 1034/IC963U82  
; CURRENT APPLICATION NUMBER: US/09/950,084  
; CURRENT FILING DATE: 2001-09-10  
; PRIOR APPLICATION NUMBER: US 09/417,811  
; PRIOR FILING DATE: 1999-10-14  
; PRIOR APPLICATION NUMBER: US 09/353,718  
; PRIOR FILING DATE: 1999-07-14  
; PRIOR APPLICATION NUMBER: US 09/266,557  
; PRIOR FILING DATE: 1999-03-11  
; PRIOR APPLICATION NUMBER: US 09/266,556  
; PRIOR FILING DATE: 1999-03-11  
; PRIOR APPLICATION NUMBER: US 09/266,555  
; PRIOR FILING DATE: 1999-03-11  
; PRIOR APPLICATION NUMBER: US 09/266,542  
; PRIOR FILING DATE: 1999-03-11  
; PRIOR APPLICATION NUMBER: US 09/266,541  
; PRIOR FILING DATE: 1999-03-11  
; PRIOR APPLICATION NUMBER: US 09/037,934  
; PRIOR FILING DATE: 1998-03-10  
; PRIOR APPLICATION NUMBER: US 09/036,720  
; PRIOR FILING DATE: 1998-03-06  
; PRIOR APPLICATION NUMBER: US 09/036,338  
; PRIOR FILING DATE: 1998-03-06  
; Remaining Prior Application data removed - See File Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 7451  
; SEQ ID NO 6523  
; LENGTH: 129  
; TYPE: PRT  
; ORGANISM: Staphylococcus aureus  
US-09-950-084-6523

Query Match 100.0%; Score 574; DB 25; Length 129;  
Best Local Similarity 100.0%; Pred. No. 4.2e-49;  
Matches 115; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MSKINDINDLVNATFQVKFFRDTKKFNLYNEEIVILNHLRSENEISSKEIAKCSF 60  
Db 15 MSKINDINDLVNATFQVKFFRDTKKFNLYNEEIVILNHLRSENEISSKEIAKCSF 74  
Qy 61 KPYLTAKLQKLDKLLSKRSLODERTVIVVYDTQKANTQKLISELEYIKN 115  
Db 75 KPYLTAKLQKLDKLLSKRSLODERTVIVVYDTQKANTQKLISELEYIKN 129  
|||||

RESULT 7  
US-09-450-969-6487  
; Sequence 6487, Application US/09450969  
; GENERAL INFORMATION:  
; APPLICANT: Lynn Doucette-Stamm et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
; FILE REFERENCE: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS  
; CURRENT APPLICATION NUMBER: US/09/450,969  
; CURRENT FILING DATE: 1999-11-29  
; NUMBER OF SEQ ID NOS: 7544

; SEQ ID NO 6487  
; LENGTH: 128  
; TYPE: PRT  
; ORGANISM: S.epidermidis  
US-09-450-969-6487

Query Match 84.7%; Score 486; DB 18; Length 128;  
Best Local Similarity 84.2%; Pred. No. 2.7e-40;  
Matches 96; Conservative 8; Mismatches 10; Indels 0; Gaps 0;  
QY 1 MSKINDINDLVNATFQVKKFFRDTKKKFNLYVEEYIILNHILRSNESISKEIAKGCSEF 60  
DB 15 MGKIKDINDLVNATFQVKKFFRDTKKQNLNVEEYIILNHILKSSESISKEIATCSEF 74  
QY 61 KPYILTALQKLKDLKLSKRSLODERTVIVYVTDTKANIQLISELEBYIK 114  
DB 75 KPYILTALQKLKDLNLLSKRSVHDERTVIVFVSDEQREKIKKLIILELNYIK 128

## RESULT 8

US-10-092-411A-4993  
; Sequence 4993, Application US/10092411A  
; GENERAL INFORMATION:  
; APPLICANT: Lynn Doucette-Stamm et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 032796-101  
; CURRENT APPLICATION NUMBER: US/10/092,411A  
; CURRENT FILING DATE: 2002-03-07  
; PRIOR APPLICATION NUMBER: US 09/134,001  
; PRIOR FILING DATE: 1998-08-13  
; PRIOR APPLICATION NUMBER: US 60/064,964  
; PRIOR FILING DATE: 1997-11-08  
; PRIOR APPLICATION NUMBER: US 60/055,779  
; PRIOR FILING DATE: 1997-08-14  
; NUMBER OF SEQ ID NOS: 5676  
; SEQ ID NO 4993  
; LENGTH: 128  
; TYPE: PRT  
; ORGANISM: Staphylococcus epidermidis  
US-10-092-411A-4993

Query Match 84.7%; Score 486; DB 26; Length 128;  
Best Local Similarity 84.2%; Pred. No. 2.7e-40;  
Matches 96; Conservative 8; Mismatches 10; Indels 0; Gaps 0;  
QY 1 MSKINDINDLVNATFQVKKFFRDTKKKFNLYVEEYIILNHILRSNESISKEIAKGCSEF 60  
DB 15 MGKIKDINDLVNATFQVKKFFRDTKKQNLNVEEYIILNHILKSSESISKEIATCSEF 74  
QY 61 KPYILTALQKLKDLKLSKRSLODERTVIVYVTDTKANIQLISELEBYIK 114  
DB 75 KPYILTALQKLKDLNLLSKRSVHDERTVIVFVSDEQREKIKKLIILELNYIK 128

## RESULT 9

PCT-US02-00877-26  
; Sequence 26, Application PC/TUS0200877  
; GENERAL INFORMATION:  
; APPLICANT: Trustees of Dartmouth College  
; TITLE OF INVENTION: Compositions and Methods for affecting virulence determinants in  
; TITLE OF INVENTION: bacteria  
; FILE REFERENCE: 11312-006-228  
; CURRENT APPLICATION NUMBER: PCT/US02/00877  
; CURRENT FILING DATE: 2003-03-04  
; PRIOR APPLICATION NUMBER: 60/261,233  
; PRIOR FILING DATE: 2001-01-12  
; PRIOR APPLICATION NUMBER: 60/261,607  
; PRIOR FILING DATE: 2001-01-12  
; PRIOR APPLICATION NUMBER: 60/289,601  
; PRIOR FILING DATE: 2001-05-08  
; NUMBER OF SEQ ID NOS: 35  
; SOFTWARE: PatentIn version 3.2

; SEQ ID NO 26  
; LENGTH: 120  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: Consensus  
; FEATURE:  
; NAME/KEY: SITE  
; LOCATION: 2,4,10,15,16,17,19,20,24,25,31,38,44,45,47,51,52,58,59,61,73,77,78,86,  
; LOCATION: 87,95,97,99,102, 106,110,114,117,118,120  
; OTHER INFORMATION: Xaa = any amino acid  
PCT-US02-00877-26

Query Match 39.0%; Score 224; DB 1; Length 120;  
Best Local Similarity 43.9%; Pred. No. 4.2e-14;  
Matches 50; Conservative 15; Mismatches 49; Indels 0; Gaps 0;  
QY 1 MSKINDINDLVNATFQVKKFFRDTKKKFNLYVEEYIILNHILRSNESISKEIAKGCSEF 60  
DB 3 MXKINDIXDLLNKKXXKFKXKIKKPEKXLSPEEFILTYIXXQXENEXLKDIIIXLXY 62  
QY 61 KPYILTALQKLKDLKLSKRSLODERTVIVYVTDTKANIQLISELEBYIK 114  
DB 63 KOPQLVKALKLKKXXYLSKGSXXDERTVLIXVXDQKXKIELLSXVNOXIK 116

## RESULT 10

US-10-043-539-26  
; Sequence 26, Application US/10043539  
; GENERAL INFORMATION:  
; APPLICANT: Cheung, Ambrose  
; APPLICANT: Manna, Adar  
; APPLICANT: Zhang, Gongyi  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR AFFECTING VIRULENCE DETERMINANTS IN  
; TITLE OF INVENTION: BACTERIA  
; FILE REFERENCE: DC-0199  
; CURRENT APPLICATION NUMBER: US/10/043,539  
; CURRENT FILING DATE: 2002-01-11  
; PRIOR APPLICATION NUMBER: US 60/261,233  
; PRIOR FILING DATE: 2001-01-12  
; PRIOR APPLICATION NUMBER: US 60/261,607  
; PRIOR FILING DATE: 2001-01-12  
; PRIOR APPLICATION NUMBER: US 60/289,601  
; PRIOR FILING DATE: 2001-05-08  
; NUMBER OF SEQ ID NOS: 34  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 26  
; LENGTH: 120  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Consensus sequence.  
; FEATURE:  
; NAME/KEY: MISC FEATURE  
; LOCATION: (2)..(2)  
; OTHER INFORMATION: "X" is defined as any amino acid residue.  
; FEATURE:  
; NAME/KEY: MISC FEATURE  
; LOCATION: (4)..(4)  
; OTHER INFORMATION: "X" is defined as any amino acid residue.  
; FEATURE:  
; NAME/KEY: MISC FEATURE  
; LOCATION: (10)..(10)  
; OTHER INFORMATION: "X" is defined as any amino acid residue.  
; FEATURE:  
; NAME/KEY: MISC FEATURE  
; LOCATION: (15)..(17)  
; OTHER INFORMATION: "X" is defined as any amino acid residue.  
; FEATURE:  
; NAME/KEY: MISC FEATURE  
; LOCATION: (19)..(20)  
; OTHER INFORMATION: "X" is defined as any amino acid residue.  
; FEATURE:

```

; NAME/KEY: MISC_FEATURE
; LOCATION: (24)..(25)
; OTHER INFORMATION: "X" is defined as any amino acid residue.
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (31)..(31)
; OTHER INFORMATION: "X" is defined as any amino acid residue.
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (38)..(38)
; OTHER INFORMATION: "X" is defined as any amino acid residue.
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (44)..(45)
; OTHER INFORMATION: "X" is defined as any amino acid residue.
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (47)..(47)
; OTHER INFORMATION: "X" is defined as any amino acid residue.
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (51)..(52)
; OTHER INFORMATION: "X" is defined as any amino acid residue.
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (58)..(59)
; OTHER INFORMATION: "X" is defined as any amino acid residue.
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (61)..(61)
; OTHER INFORMATION: "X" is defined as any amino acid residue.
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (73)..(73)
; OTHER INFORMATION: "X" is defined as any amino acid residue.
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (77)..(78)
; OTHER INFORMATION: "X" is defined as any amino acid residue.
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (86)..(87)
; OTHER INFORMATION: "X" is defined as any amino acid residue.
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (95)..(95)
; OTHER INFORMATION: "X" is defined as any amino acid residue.
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (97)..(97)
; OTHER INFORMATION: "X" is defined as any amino acid residue.
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (99)..(99)
; OTHER INFORMATION: "X" is defined as any amino acid residue.
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (102)..(102)
; OTHER INFORMATION: "X" is defined as any amino acid residue.
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (106)..(106)
; OTHER INFORMATION: "X" is defined as any amino acid residue.
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (110)..(110)
; OTHER INFORMATION: "X" is defined as any amino acid residue.
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (114)..(114)
; OTHER INFORMATION: "X" is defined as any amino acid residue.
; FEATURE:
; NAME/KEY: MISC_FEATURE

```

```

; LOCATION: (117)..(118)
; OTHER INFORMATION: "X" is defined as any amino acid residue.
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (120)..(120)
; OTHER INFORMATION: "X" is defined as any amino acid residue.
; US-10-043-539-26

Query Match      39.0%; Score 224; DB 26; Length 120;
Best Local Similarity 43.9%; Pred. No. 4.2e-14;
Matches 50; Conservative 15; Mismatches 49; Indels 0; Gaps 0;

Qy 1 MSKINDINDLVNATFQVKPFRTDKKENLVNVEIYILAHILRSENEISSKRIACSEF 60
Db 3 MXKINDIXDLNXXYYXKFKXXIKKFKXLSFEEFXYLYXXQXENEXXKDLIXLXY 62

Qy 61 KPYVLTKALQKLDKLLSKRSLOBERTVIVVYDTQKANIOLKLISELEYIK 114
Db 63 KQQLVKALKXLLKXXYYLSKRSXDXBERTVLIXVXDQXKXIEKXLLSXVNXQIK 116

RESULT 11
US-10-043-539A-26
; Sequence 26, Application US/10043539A
; GENERAL INFORMATION:
; APPLICANT: Cheung, Ambrose
; APPLICANT: Manna, Adar
; APPLICANT: Zhang, Gongyi
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR AFFECTING VIRULENCE DETERMINANTS IN
; FILE REFERENCE: DC-0199
; CURRENT APPLICATION NUMBER: US/10/043,539A
; CURRENT FILING DATE: 2000-01-11
; PRIOR APPLICATION NUMBER: US 60/261,233
; PRIOR FILING DATE: 2001-01-12
; PRIOR APPLICATION NUMBER: US 60/261,607
; PRIOR FILING DATE: 2001-01-12
; PRIOR APPLICATION NUMBER: US 60/289,601
; PRIOR FILING DATE: 2001-05-08
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 26
; LENGTH: 120
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Consensus sequence.
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (2)..(2)
; OTHER INFORMATION: "X" is defined as any amino acid residue.
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (4)..(4)
; OTHER INFORMATION: "X" is defined as any amino acid residue.
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (10)..(10)
; OTHER INFORMATION: "X" is defined as any amino acid residue.
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (15)..(17)
; OTHER INFORMATION: "X" is defined as any amino acid residue.
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (19)..(20)
; OTHER INFORMATION: "X" is defined as any amino acid residue.
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (24)..(25)
; OTHER INFORMATION: "X" is defined as any amino acid residue.
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (31)..(31)

```

OTHER INFORMATION: "X" is defined as any amino acid residue.  
FEATURE:  
NAME/KEY: MISC\_FEATURE  
LOCATION: (38)..(38)  
OTHER INFORMATION: "X" is defined as any amino acid residue.  
FEATURE:  
NAME/KEY: MISC\_FEATURE  
LOCATION: (44)..(45)  
OTHER INFORMATION: "X" is defined as any amino acid residue.  
FEATURE:  
NAME/KEY: MISC\_FEATURE  
LOCATION: (47)..(47)  
OTHER INFORMATION: "X" is defined as any amino acid residue.  
FEATURE:  
NAME/KEY: MISC\_FEATURE  
LOCATION: (51)..(52)  
OTHER INFORMATION: "X" is defined as any amino acid residue.  
FEATURE:  
NAME/KEY: MISC\_FEATURE  
LOCATION: (58)..(59)  
OTHER INFORMATION: "X" is defined as any amino acid residue.  
FEATURE:  
NAME/KEY: MISC\_FEATURE  
LOCATION: (61)..(61)  
OTHER INFORMATION: "X" is defined as any amino acid residue.  
FEATURE:  
NAME/KEY: MISC\_FEATURE  
LOCATION: (73)..(73)  
OTHER INFORMATION: "X" is defined as any amino acid residue.  
FEATURE:  
NAME/KEY: MISC\_FEATURE  
LOCATION: (77)..(78)  
OTHER INFORMATION: "X" is defined as any amino acid residue.  
FEATURE:  
NAME/KEY: MISC\_FEATURE  
LOCATION: (86)..(87)  
OTHER INFORMATION: "X" is defined as any amino acid residue.  
FEATURE:  
NAME/KEY: MISC\_FEATURE  
LOCATION: (95)..(95)  
OTHER INFORMATION: "X" is defined as any amino acid residue.  
FEATURE:  
NAME/KEY: MISC\_FEATURE  
LOCATION: (97)..(97)  
OTHER INFORMATION: "X" is defined as any amino acid residue.  
FEATURE:  
NAME/KEY: MISC\_FEATURE  
LOCATION: (99)..(99)  
OTHER INFORMATION: "X" is defined as any amino acid residue.  
FEATURE:  
NAME/KEY: MISC\_FEATURE  
LOCATION: (102)..(102)  
OTHER INFORMATION: "X" is defined as any amino acid residue.  
FEATURE:  
NAME/KEY: MISC\_FEATURE  
LOCATION: (106)..(106)  
OTHER INFORMATION: "X" is defined as any amino acid residue.  
FEATURE:  
NAME/KEY: MISC\_FEATURE  
LOCATION: (110)..(110)  
OTHER INFORMATION: "X" is defined as any amino acid residue.  
FEATURE:  
NAME/KEY: MISC\_FEATURE  
LOCATION: (114)..(114)  
OTHER INFORMATION: "X" is defined as any amino acid residue.  
FEATURE:  
NAME/KEY: MISC\_FEATURE  
LOCATION: (117)..(118)  
OTHER INFORMATION: "X" is defined as any amino acid residue.  
FEATURE:  
NAME/KEY: MISC\_FEATURE  
LOCATION: (120)..(120)  
OTHER INFORMATION: "X" is defined as any amino acid residue.

US-10-043-539A-26  
Query Match 39.0%; Score 224; DB 26; Length 120;  
Best Local Similarity 43.9%; Pred. No. 4.2e-14;  
Matches 50; Conservative 15; Mismatches 49; Indels 0; Gaps 0;  
QY 1 MSKINDINDLVNATFQVKFFRDTKKFNLNVEEIVVYVDTOKANIQLISELEEVK 114  
Db 3 MKKINDIXDLLNKKXXKFKXIKKEFXLSPEEFXILTYIXXQXENEXLKDIIIXLXY 62  
QY 61 KPYLTALQKLKDLKLSKESLODERTVIVVYVDTOKANIQLISELEEVK 114  
Db 63 KOPQLVKALKXLLKXXYLSKRSXDXDERTVLIXVXDXORXKIELLSVNXQIK 116  
RESULT 12  
PCT-US02-00877-35  
; Sequence 35, Application PC/TUS0200877  
; GENERAL INFORMATION:  
; APPLICANT: Trustees of Dartmouth College  
; TITLE OF INVENTION: Compositions and Methods for affecting virulence determinants in  
; TITLE OF INVENTION: Bacteria  
; FILE REFERENCE: 11312-006-228  
; CURRENT APPLICATION NUMBER: PCT/US02/00877  
; CURRENT FILING DATE: 2003-03-04  
; PRIOR APPLICATION NUMBER: 60/261,233  
; PRIOR FILING DATE: 2001-01-12  
; PRIOR APPLICATION NUMBER: 60/261,607  
; PRIOR FILING DATE: 2001-01-12  
; PRIOR APPLICATION NUMBER: 60/289,601  
; PRIOR FILING DATE: 2001-05-08  
; NUMBER OF SEQ ID NOS: 35  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 35  
; LENGTH: 120  
; TYPE: PRT  
; ORGANISM: Staphylococcus aureus  
PCT-US02-00877-35  
Query Match 25.8%; Score 148; DB 1; Length 120;  
Best Local Similarity 30.1%; Pred. No. 1.7e-06;  
Matches 34; Conservative 27; Mismatches 52; Indels 0; Gaps 0;  
QY 1 MSKINDINDLVNATFQVKFFRDTKKFNLNVEEIVVYVDTOKANIQLISELEEVK 60  
Db 3 ITKINDCFELLSWVTVYADKLKSLIKKPSISFEFAVLTYISENKEVEYLKDIINHLNY 62  
QY 61 KPYLTALQKLKDLKLSKESLODERTVIVVYVDTOKANIQLISELEEVK 113  
Db 63 KQPQVVKAVKILSQSDYFDKKGNEHDERTVLILVNAQQRKKIESLLSRVTKRI 115  
RESULT 13  
US-10-043-539A-35  
; Sequence 35, Application US/10043539A  
; GENERAL INFORMATION:  
; APPLICANT: Cheung, Ambrose  
; APPLICANT: Manna, Adar  
; APPLICANT: Zhang, Gongyi  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR AFFECTING VIRULENCE DETERMINANTS IN E  
; FILE REFERENCE: DC-0199  
; CURRENT APPLICATION NUMBER: US/10/043,539A  
; CURRENT FILING DATE: 2000-01-11  
; PRIOR APPLICATION NUMBER: US 60/261,233  
; PRIOR FILING DATE: 2001-01-12  
; PRIOR APPLICATION NUMBER: US 60/261,607  
; PRIOR FILING DATE: 2001-01-12  
; PRIOR APPLICATION NUMBER: US 60/289,601  
; PRIOR FILING DATE: 2001-05-08  
; NUMBER OF SEQ ID NOS: 35  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 35  
; LENGTH: 120

```
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-10-043-539A-35

Query Match      25.8%; Score 148; DB 26; Length 120;
Best Local Similarity 30.1%; Pred. No. 1.7e-06;
Matches 34; Conservative 27; Mismatches 52; Indels 0; Gaps 0;

QY   1 MSKINDINDLVNATFQVKKFRTDKKKFNUNYEIYLNLHLSSENEISSKEIAKCEF 60
     :|||||:|::||::||::||::||::||::||::||::||::||::||::||:
Db    3 ITKINDCFELLSSVMVTADKLKSIIKKFEFSISPEFAVLTVISENKEEYVLDIINHLNY 62
     QY   61 KPYLTKALOKLDKLKLSKKRSLODERTVIIVVTTDTOKANIOKLISELEEI 113
     Db    63 KQPQWKAIVILSOEDYFDKKRNEHDERTVLIILVNAOQRKKIESLLSRVNKRI 115

RESULT 14
PCT-US02-00877-3
; SEQUENCE 3, Application PC/TUS0200877
; GENERAL INFORMATION:
; APPLICANT: Trustees of Dartmouth College
; TITLE OF INVENTION: Compositions and Methods for affecting virulence determinants in
; TITLE OF INVENTION: bacteria
; FILE REFERENCE: 11312-006-228
; CURRENT APPLICATION NUMBER: PCT/US02/00877
; CURRENT FILING DATE: 2003-03-04
; PRIOR APPLICATION NUMBER: 60/261,233
; PRIOR FILING DATE: 2001-01-12
; PRIOR APPLICATION NUMBER: 60/261,607
; PRIOR FILING DATE: 2001-01-12
; PRIOR APPLICATION NUMBER: 60/289,601
; PRIOR FILING DATE: 2001-05-08
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 3
; LENGTH: 124
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
PCT-US02-00877-3

Query Match      25.8%; Score 148; DB 1; Length 124;
Best Local Similarity 30.1%; Pred. No. 1.8e-06;
Matches 34; Conservative 27; Mismatches 52; Indels 0; Gaps 0;

QY   1 MSKINDINDLVNATFQVKKFRTDKKKFNUNYEIYLNLHLSSENEISSKEIAKCEF 60
     :|||||:|::||::||::||::||::||::||::||::||::||::||:
Db    3 ITKINDCFELLSSVMVTADKLKSIIKKFEFSISPEFAVLTVISENKEEYVLDIINHLNY 62
     QY   61 KPYLTKALOKLDKLKLSKKRSLODERTVIIVVTTDTOKANIOKLISELEEI 113
     Db    63 KQPQWKAIVILSOEDYFDKKRNEHDERTVLIILVNAOQRKKIESLLSRVNKRI 115

RESULT 15
US-09-612-549A-1
; SEQUENCE 1, Application US/09612549A
; GENERAL INFORMATION:
; APPLICANT: HURLBURT, Barry
; APPLICANT: SMELTZER, Mark
; APPLICANT: RECHTIN, Tammy
; TITLE OF INVENTION: INHIBITORS OF STAPHYLOCOCCUS SARA PROTEIN FUNCTION INVOLVED IN THE
; TITLE OF INVENTION: EXPRESSION OF STAPHYLOCOCCAL VIRULENCE FACTORS AND THE USE THEREOF
; FILE REFERENCE: 023533/0126
; CURRENT APPLICATION NUMBER: US/09/612,549A
; CURRENT FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: US 60/142,793
; PRIOR FILING DATE: 1999-07-08
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 124
```

**This Page Blank (uspto)**



```
; PRIOR APPLICATION NUMBER: US 60/329,140
; PRIOR FILING DATE: 2001-10-12
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 147
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-10-469-477-2

Query Match      16.3%; Score 93.5; DB 6; Length 147;
Best Local Similarity 28.4%; Pred. No. 9.3;
Matches 25; Conservative 25; Mismatches 35; Indels 3; Gaps 3;

QY 10 LVNATFQVKFFRD-TKKKFNLYEEIYLHILSRSSNEISSKEIAKCSFEPKPYLTKA 68
Db 15 LVNAQRQNRYSYKVKFNLYTPQFLVLT-ILWDES-FVNVKVVTEALDGTGVSPL 72

QY 69 LQKLKDLKLLSKRSLODERTVIVVTD 96
Db 73 LKRMEQVDLIKRSEVDQREVFHLD 100

RESULT 3
US-10-679-063-18206
; Sequence 18206, Application US/10679063
; GENERAL INFORMATION:
; APPLICANT: Edgerton, Michael D
; TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES
; FILE REFERENCE: 38-15(52054)B
; CURRENT APPLICATION NUMBER: US/10/679,063
; CURRENT FILING DATE: 2003-10-02
; PRIOR APPLICATION NUMBER: 60/415,758
; PRIOR FILING DATE: 2002-10-02
; NUMBER OF SEQ ID NOS: 27373
; SEQ ID NO 18206
; LENGTH: 513
; TYPE: PRT
; ORGANISM: Plasmodium falciparum 3D7
US-10-679-063-18206

Query Match      15.9%; Score 91; DB 6; Length 513;
Best Local Similarity 28.0%; Pred. No. 29;
Matches 28; Conservative 24; Mismatches 38; Indels 10; Gaps 4;

QY 17 VKKFRDTPKKNLYEEIYLHILRSE--SNEISSKEIAKCSFEPKPYLTKALQKL-- 72
Db 27 VKKQDDINLRKIEYNLFTQNELLRNDPLNHEKLTKEVQSLE---DANKKIQIK 82

QY 73 -KDLKLLSKRSLODERTVIVVTDTKANIKLISELEE 111
Db 83 ERDYIMHKKRVLQEKETLNKEI-HKQKHEIKTQNSVDE 121

RESULT 4
US-10-679-063-20147
; Sequence 20147, Application US/10679063
; GENERAL INFORMATION:
; APPLICANT: Edgerton, Michael D
; TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES
; FILE REFERENCE: 38-15(52054)B
; CURRENT APPLICATION NUMBER: US/10/679,063
; CURRENT FILING DATE: 2003-10-02
; PRIOR APPLICATION NUMBER: 60/415,758
; PRIOR FILING DATE: 2002-10-02
; NUMBER OF SEQ ID NOS: 27373
; SEQ ID NO 20147
; LENGTH: 1378
; TYPE: PRT
; ORGANISM: Rattus norvegicus
US-10-679-063-20147

Query Match      15.6%; Score 89.5; DB 6; Length 1378;
Best Local Similarity 24.2%; Pred. No. 65;
Matches 32; Conservative 28; Mismatches 51; Indels 21; Gaps 6;

QY 2 SKINDINDLVNATFQVKKFFRD-TKKKFNLYEE-----IYLHILSRSSNEISSKEI 54
Db 99 SKLGHVTSLSNVE--KEFILDIPNPKPMTLATEEPVPFKTLVLNQEQAETSLMRKTL 155

QY 55 AKCSFEPKPYLTK---ALQKLKDLK-----LISKKSLODERTVIVVTDTKANIQK 104
Db 156 KSAFQVQETFLNEKPTJLEKTKDYKFDSELMTSRRDKPEEATIIEMTDLKEPDGK 215

QY 105 L-ISELEEYIKN 115
Db 216 VLTSSPLYIKN 227

RESULT 5
US-10-673-119-25
; Sequence 25, Application US/10673119
; GENERAL INFORMATION:
; APPLICANT: O'Donnell, Michael E.
; APPLICANT: Yuzhakov, Alexander
; APPLICANT: Yurieva, Olga
; APPLICANT: Jeruzalmi, David
; APPLICANT: Bruck, Irina
; APPLICANT: Kuriyan, John
; TITLE OF INVENTION: ENZYMES DERIVED FROM THERMOPHILIC ORGANISMS THAT
; TITLE OF INVENTION: FUNCTION AS A CHROMOSOMAL REPLICASE, PREPARATION AND
; FILE REFERENCE: 22221/1030
; CURRENT APPLICATION NUMBER: US/10/673,119
; CURRENT FILING DATE: 2003-09-26
; PRIOR APPLICATION NUMBER: US/09/716,964A
; PRIOR FILING DATE: 2000-11-21
; PRIOR APPLICATION NUMBER: 60/143,202
; PRIOR FILING DATE: 1997-04-08
; PRIOR APPLICATION NUMBER: 08/823,407
; PRIOR FILING DATE: 1997-04-08
; PRIOR APPLICATION NUMBER: 09/057,416
; PRIOR FILING DATE: 1998-04-08
; NUMBER OF SEQ ID NOS: 212
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 25
; LENGTH: 260
; TYPE: PRT
; ORGANISM: Mycoplasma genitalium
US-10-673-119-25

Query Match      14.2%; Score 81.5; DB 6; Length 260;
Best Local Similarity 22.8%; Pred. No. 70;
Matches 31; Conservative 27; Mismatches 37; Indels 41; Gaps 6;

QY 4 INDINDLVNATFQVKKFFRDTPKKNLYEEIYLN--HILRSE----- 46
Db 100 INDIRELVENFVNHPTF--KKK-----VYLDEAHMLTTQSGGLLTLESPPYV 149

QY 47 -----NEISSKEIAKCS--FKPYLTKALQKLKDLKLLSKRSLODERTVIVVVT 95
Db 150 LRFITTFENKIPLTILSRCSQSFKKITSLLERLND---IAKKEIKEDALINIA 206

QY 96 DTQKANIQLISELEE 111
Db 207 DLSQSLRDGLSLDQ 222

RESULT 6
US-10-673-120-25
; Sequence 25, Application US/10673120
; GENERAL INFORMATION:
; APPLICANT: O'Donnell, Michael E.
; APPLICANT: Yuzhakov, Alexander
; APPLICANT: Yurieva, Olga
; APPLICANT: Jeruzalmi, David
```



```

; APPLICANT: Bruck, Irina
; APPLICANT: Kuriyan, John
; TITLE OF INVENTION: ENZYMES DERIVED FROM THERMOPHILIC ORGANISMS THAT
; TITLE OF INVENTION: FUNCTION AS A CHROMOSOMAL REPLICASE, PREPARATION AND
; FILE REFERENCE: 22221/1030
; CURRENT APPLICATION NUMBER: US/10/673,120
; CURRENT FILING DATE: 2003-09-26
; PRIOR APPLICATION NUMBER: US/09/716,964A
; PRIOR FILING DATE: 2000-11-21
; PRIOR APPLICATION NUMBER: 60/143,202
; PRIOR FILING DATE: 1997-04-08
; PRIOR APPLICATION NUMBER: 08/823,407
; PRIOR FILING DATE: 1997-04-08
; PRIOR APPLICATION NUMBER: 09/057,416
; PRIOR FILING DATE: 1998-04-08
; NUMBER OF SEQ ID NOS: 212
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 25
; LENGTH: 260
; TYPE: PRT
; ORGANISM: Mycoplasma genitalium
US-10-673-120-25

Query Match      14.2%; Score 81.5; DB 6; Length 260;
Best Local Similarity 22.8%; Pred. No. 70;
Matches 31; Conservative 27; Mismatches 37; Indels 41; Gaps 6;

QY 4 INDINDLVNATFQVKKFFRDTKKFNLYEIIYLN--HILRSES-----46
Db 100 INDIRELVENVNHPFTF---KKK-----VYLDEAHLTTQSMGGLKLTLESPYV 149
QY 47 -----NEISSKEIAKSE--FKPYLTALQKLDKLLSKRSLODERTVIVYVT 95
Db 150 LFIPTTFENKIPLTILSRQCSFFFKITSDILERLND---IAKKEKIKEDALIKIA 206
QY 96 DTQKANIQKLISEEE 111
Db 207 DLSQGSRLDGLSLDQ 222

RESULT 7
US-10-671-134-25
; Sequence 25, Application US/10671134
; GENERAL INFORMATION:
; APPLICANT: O'Donnell, Michael E.
; APPLICANT: Yuzhakov, Alexander
; APPLICANT: Yurieva, Olga
; APPLICANT: Jeruzalmi, David
; APPLICANT: Bruck, Irina
; APPLICANT: Kuriyan, John
; TITLE OF INVENTION: ENZYMES DERIVED FROM THERMOPHILIC ORGANISMS THAT
; TITLE OF INVENTION: FUNCTION AS A CHROMOSOMAL REPLICASE, PREPARATION AND
; FILE REFERENCE: 22221/1030
; CURRENT APPLICATION NUMBER: US/10/671,134
; CURRENT FILING DATE: 2003-09-25
; PRIOR APPLICATION NUMBER: 60/143,202
; PRIOR FILING DATE: 1997-04-08
; PRIOR APPLICATION NUMBER: 08/823,407
; PRIOR FILING DATE: 1997-04-08
; PRIOR APPLICATION NUMBER: 09/057,416
; PRIOR FILING DATE: 1998-04-08
; NUMBER OF SEQ ID NOS: 212
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 25
; LENGTH: 260
; TYPE: PRT
; ORGANISM: Mycoplasma genitalium
US-10-671-134-25

Query Match      14.2%; Score 81.5; DB 6; Length 260;
Best Local Similarity 22.8%; Pred. No. 70;
Matches 31; Conservative 27; Mismatches 37; Indels 41; Gaps 6;

QY 4 INDINDLVNATFQVKKFFRDTKKFNLYEIIYLN--HILRSES-----46
Db 100 INDIRELVENVNHPFTF---KKK-----VYLDEAHLTTQSMGGLKLTLESPYV 149
QY 47 -----NEISSKEIAKSE--FKPYLTALQKLDKLLSKRSLODERTVIVYVT 95
Db 150 LFIPTTFENKIPLTILSRQCSFFFKITSDILERLND---IAKKEKIKEDALIKIA 206
QY 96 DTQKANIQKLISEEE 111
Db 207 DLSQGSRLDGLSLDQ 222

RESULT 8
US-10-673-098-25
; Sequence 25, Application US/10673098
; GENERAL INFORMATION:
; APPLICANT: O'Donnell, Michael E.
; APPLICANT: Yuzhakov, Alexander
; APPLICANT: Yurieva, Olga
; APPLICANT: Jeruzalmi, David
; APPLICANT: Bruck, Irina
; APPLICANT: Kuriyan, John
; TITLE OF INVENTION: ENZYMES DERIVED FROM THERMOPHILIC ORGANISMS THAT
; TITLE OF INVENTION: FUNCTION AS A CHROMOSOMAL REPLICASE, PREPARATION AND
; FILE REFERENCE: 22221/1030
; CURRENT APPLICATION NUMBER: US/10/673,098
; CURRENT FILING DATE: 2003-09-26
; PRIOR APPLICATION NUMBER: US/09/716,964A
; PRIOR FILING DATE: 2000-11-21
; PRIOR APPLICATION NUMBER: 60/143,202
; PRIOR FILING DATE: 1997-04-08
; PRIOR APPLICATION NUMBER: 08/823,407
; PRIOR FILING DATE: 1997-04-08
; PRIOR APPLICATION NUMBER: 09/057,416
; PRIOR FILING DATE: 1998-04-08
; NUMBER OF SEQ ID NOS: 212
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 25
; LENGTH: 260
; TYPE: PRT
; ORGANISM: Mycoplasma genitalium
US-10-673-098-25

Query Match      14.2%; Score 81.5; DB 6; Length 260;
Best Local Similarity 22.8%; Pred. No. 70;
Matches 31; Conservative 27; Mismatches 37; Indels 41; Gaps 6;

QY 4 INDINDLVNATFQVKKFFRDTKKFNLYEIIYLN--HILRSES-----46
Db 100 INDIRELVENVNHPFTF---KKK-----VYLDEAHLTTQSMGGLKLTLESPYV 149
QY 47 -----NEISSKEIAKSE--FKPYLTALQKLDKLLSKRSLODERTVIVYVT 95
Db 150 LFIPTTFENKIPLTILSRQCSFFFKITSDILERLND---IAKKEKIKEDALIKIA 206
QY 96 DTQKANIQKLISEEE 111
Db 207 DLSQGSRLDGLSLDQ 222

RESULT 9
US-10-673-127-25
; Sequence 25, Application US/10673127
; GENERAL INFORMATION:
; APPLICANT: O'Donnell, Michael E.
; APPLICANT: Yuzhakov, Alexander
; APPLICANT: Yurieva, Olga

```

```

; APPLICANT: Jeruzalmi, David
; APPLICANT: Bruck, Irina
; APPLICANT: Kuriyan, John
; TITLE OF INVENTION: ENZYMES DERIVED FROM THERMOPHILIC ORGANISMS THAT
; TITLE OF INVENTION: FUNCTION AS A CHROMOSOMAL REPLICASE, PREPARATION AND
; TITLE OF INVENTION: USE THEREOF
; FILE REFERENCE: 2221/1030
; CURRENT APPLICATION NUMBER: US/10/673,127
; CURRENT FILING DATE: 2003-09-26
; PRIOR APPLICATION NUMBER: US/09/716,964A
; PRIOR FILING DATE: 2000-11-21
; PRIOR APPLICATION NUMBER: 60/143,202
; PRIOR FILING DATE: 1997-04-08
; PRIOR APPLICATION NUMBER: 08/823,407
; PRIOR FILING DATE: 1997-04-08
; PRIOR APPLICATION NUMBER: 09/057,416
; PRIOR FILING DATE: 1998-04-08
; NUMBER OF SEQ ID NOS: 212
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 25
; LENGTH: 260
; TYPE: PRT
; ORGANISM: Mycoplasma genitalium
; US-10-673-127-25

Query Match 14.2%; Score 81.5; DB 6; Length 260;
Best Local Similarity 22.8%; Pred No. 70;
Matches 31; Conservative 27; Mismatches 37; Indels 41; Gaps 6;

QY 4 INDINDLVNATFQVKKFFRTKKNFNLYEIIYILN--HILRSSE----- 46
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 100 INDIRELVENVFNHPTTF--KKK-----VVILDEAHMLTQSWGGLKLTLEESPPV 149

QY 47 -----NIISSKETAKCSE--FKPYIYTKLQKLKLLSKKRSIQDERTVIVVT 95
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 150 LFITTFTEFNKIPITLSRCQSFFFKITSDLILERLND---IAKKEKIKIEKDALKIA 206

QY 96 DTQKANIQKLISELEE 111
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 207 DLSQGSRLDGLSLDQ 222

RESULT 10
US-10-670-844-25
; Sequence 25, Application US/10670844
; GENERAL INFORMATION:
; APPLICANT: O'Donnell, Michael E.
; APPLICANT: Yuzhakov, Alexander
; APPLICANT: Yurieva, Olga
; APPLICANT: Jeruzalmi, David
; APPLICANT: Bruck, Irina
; APPLICANT: Kuriyan, John
; TITLE OF INVENTION: ENZYMES DERIVED FROM THERMOPHILIC ORGANISMS THAT
; TITLE OF INVENTION: FUNCTION AS A CHROMOSOMAL REPLICASE, PREPARATION AND
; TITLE OF INVENTION: USE THEREOF
; FILE REFERENCE: 2221/1030
; CURRENT APPLICATION NUMBER: US/10/670,844
; CURRENT FILING DATE: 2003-09-25
; PRIOR APPLICATION NUMBER: US/09/716,964A
; PRIOR FILING DATE: 2000-11-21
; PRIOR APPLICATION NUMBER: 60/143,202
; PRIOR FILING DATE: 1997-04-08
; PRIOR APPLICATION NUMBER: 08/823,407
; PRIOR FILING DATE: 1997-04-08
; PRIOR APPLICATION NUMBER: 09/057,416
; PRIOR FILING DATE: 1998-04-08
; NUMBER OF SEQ ID NOS: 212
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 25
; LENGTH: 260
; TYPE: PRT
; ORGANISM: Mycoplasma genitalium
; US-10-670-844-25

```

```

US-10-671-412-25
Query Match      14.2%; Score 81.5; DB 6; Length 260;
Best Local Similarity 22.8%; Pred. No. 70;
Matches 31; Conservative 27; Mismatches 37; Indels 41; Gaps 6;

QY 4 INDINDLVNATQVKKFFRDTKKKNLNYEIIYILN--HILRSBS-----46
DB 100 INDIRELVNENHPPTF---KKK-----VYLDEAHLMTTQSGGLLTKLEESPPYV 149
QY 47 -----NEISSKEIAKCSB--FKPYLTKALQKLDKLLSKRSLODERTVIVYVT 95
DB 150 LFIPTTTEBNKIPILSRCSQSFPPKITSLLILRLND---IAKKEKIKIEKDALIKIA 206
QY 96 DTQKANIQLISELEE 111
DB 207 DLSQGSRLDGLSLDDQ 222

RESULT 14
US-10-688-058-60
; Sequence 60, Application US/10688058
; GENERAL INFORMATION:
; APPLICANT: SYKES, KATHRYN F.
; APPLICANT: HALE, KATHERINE S.
; APPLICANT: JOHNSTON, STEPHEN A.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR VACCINATION COMPRISING
; TITLE OF INVENTION: NUCLEIC ACID AND/OR POLYPEPTIDE SEQUENCES OF THE GENUS
; TITLE OF INVENTION: BORRELIA
; FILE REFERENCE: MCRO:003US
; CURRENT APPLICATION NUMBER: US/10/688,058
; CURRENT FILING DATE: 2003-10-17
; NUMBER OF SEQ ID NOS: 141
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 60
; LENGTH: 1125
; TYPE: PRT
; ORGANISM: Borrelia burgdorferi
US-10-688-058-60

Query Match      13.9%; Score 80; DB 6; Length 1125;
Best Local Similarity 22.2%; Pred. No. 2.1e-02;
Matches 30; Conservative 27; Mismatches 50; Indels 28; Gaps 5;

QY 1 MSKINDINDLVNATQVKKFFRDT-----KKKFNLYE-----EIVILNHLR 43
DB 127 LSKIPDKNTLLKNIYKIEK---NTNINTADIEKTLITLGYEKLTVIIPGFTVKGEIID 183
QY 44 ----SESNEI-----SSKEIAKCSBPKPYLTKALQKLDKLLSKRSLODERTVIVYVT 95
DB 184 IYFPGEQNPRIALNFDKIEBKFNPLTQLKHDNEILEFQILPKKSIINDDKTINTLKT 243
QY 96 DTQKANIQLISELEE 110
DB 244 KIKSVEYKKILEELD 258

RESULT 15
PCT-US03-27401-259
; Sequence 259, Application PC/TUS0327401
; GENERAL INFORMATION:
; APPLICANT: TUFTS UNIVERSITY
; TITLE OF INVENTION: STREPTOCOCCUS PNEUMONIAE FOR DIAGNOSIS, TREATMENT AND
; TITLE OF INVENTION: PREVENTION OF ACTIVE INFECTION
; FILE REFERENCE: 700355-52941-PCT
; CURRENT APPLICATION NUMBER: PCT/US03/27401
; CURRENT FILING DATE: 2003-09-02
; PRIOR APPLICATION NUMBER: US 60/407,082
; PRIOR FILING DATE: 2002-08-30
; NUMBER OF SEQ ID NOS: 560
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 259
; LENGTH: 428

```

```

US-10-671-412-25
Query Match      14.2%; Score 81.5; DB 6; Length 260;
Best Local Similarity 22.8%; Pred. No. 70;
Matches 31; Conservative 27; Mismatches 37; Indels 41; Gaps 6;

QY 4 INDINDLVNATQVKKFFRDTKKKNLNYEIIYILN--HILRSBS-----46
DB 100 INDIRELVNENHPPTF---KKK-----VYLDEAHLMTTQSGGLLTKLEESPPYV 149
QY 47 -----NEISSKEIAKCSB--FKPYLTKALQKLDKLLSKRSLODERTVIVYVT 95
DB 150 LFIPTTTEBNKIPILSRCSQSFPPKITSLLILRLND---IAKKEKIKIEKDALIKIA 206
QY 96 DTQKANIQLISELEE 111
DB 207 DLSQGSRLDGLSLDDQ 222

RESULT 14
US-10-688-058-60
; Sequence 60, Application US/10688058
; GENERAL INFORMATION:
; APPLICANT: SYKES, KATHRYN F.
; APPLICANT: HALE, KATHERINE S.
; APPLICANT: JOHNSTON, STEPHEN A.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR VACCINATION COMPRISING
; TITLE OF INVENTION: NUCLEIC ACID AND/OR POLYPEPTIDE SEQUENCES OF THE GENUS
; TITLE OF INVENTION: BORRELIA
; FILE REFERENCE: MCRO:003US
; CURRENT APPLICATION NUMBER: US/10/688,058
; CURRENT FILING DATE: 2003-10-17
; NUMBER OF SEQ ID NOS: 141
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 60
; LENGTH: 1125
; TYPE: PRT
; ORGANISM: Borrelia burgdorferi
US-10-688-058-60

Query Match      13.9%; Score 80; DB 6; Length 1125;
Best Local Similarity 22.2%; Pred. No. 2.1e-02;
Matches 30; Conservative 27; Mismatches 50; Indels 28; Gaps 5;

QY 1 MSKINDINDLVNATQVKKFFRDT-----KKKFNLYE-----EIVILNHLR 43
DB 127 LSKIPDKNTLLKNIYKIEK---NTNINTADIEKTLITLGYEKLTVIIPGFTVKGEIID 183
QY 44 ----SESNEI-----SSKEIAKCSBPKPYLTKALQKLDKLLSKRSLODERTVIVYVT 95
DB 184 IYFPGEQNPRIALNFDKIEBKFNPLTQLKHDNEILEFQILPKKSIINDDKTINTLKT 243
QY 96 DTQKANIQLISELEE 110
DB 244 KIKSVEYKKILEELD 258

RESULT 15
PCT-US03-27401-259
; Sequence 259, Application PC/TUS0327401
; GENERAL INFORMATION:
; APPLICANT: TUFTS UNIVERSITY
; TITLE OF INVENTION: STREPTOCOCCUS PNEUMONIAE FOR DIAGNOSIS, TREATMENT AND
; TITLE OF INVENTION: PREVENTION OF ACTIVE INFECTION
; FILE REFERENCE: 700355-52941-PCT
; CURRENT APPLICATION NUMBER: PCT/US03/27401
; CURRENT FILING DATE: 2003-09-02
; PRIOR APPLICATION NUMBER: US 60/407,082
; PRIOR FILING DATE: 2002-08-30
; NUMBER OF SEQ ID NOS: 560
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 259
; LENGTH: 428

US-10-671-412-25
Query Match      14.2%; Score 81.5; DB 6; Length 260;
Best Local Similarity 22.8%; Pred. No. 70;
Matches 31; Conservative 27; Mismatches 37; Indels 41; Gaps 6;

QY 4 INDINDLVNATQVKKFFRDTKKKNLNYEIIYILN--HILRSBS-----46
DB 100 INDIRELVNENHPPTF---KKK-----VYLDEAHLMTTQSGGLLTKLEESPPYV 149
QY 47 -----NEISSKEIAKCSB--FKPYLTKALQKLDKLLSKRSLODERTVIVYVT 95
DB 150 LFIPTTTEBNKIPILSRCSQSFPPKITSLLILRLND---IAKKEKIKIEKDALIKIA 206
QY 96 DTQKANIQLISELEE 111
DB 207 DLSQGSRLDGLSLDDQ 222

RESULT 14
US-10-688-058-60
; Sequence 60, Application US/10688058
; GENERAL INFORMATION:
; APPLICANT: SYKES, KATHRYN F.
; APPLICANT: HALE, KATHERINE S.
; APPLICANT: JOHNSTON, STEPHEN A.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR VACCINATION COMPRISING
; TITLE OF INVENTION: NUCLEIC ACID AND/OR POLYPEPTIDE SEQUENCES OF THE GENUS
; TITLE OF INVENTION: BORRELIA
; FILE REFERENCE: MCRO:003US
; CURRENT APPLICATION NUMBER: US/10/688,058
; CURRENT FILING DATE: 2003-10-17
; NUMBER OF SEQ ID NOS: 141
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 60
; LENGTH: 1125
; TYPE: PRT
; ORGANISM: Borrelia burgdorferi
US-10-688-058-60

Query Match      13.9%; Score 80; DB 6; Length 1125;
Best Local Similarity 22.2%; Pred. No. 2.1e-02;
Matches 30; Conservative 27; Mismatches 50; Indels 28; Gaps 5;

QY 1 MSKINDINDLVNATQVKKFFRDT-----KKKFNLYE-----EIVILNHLR 43
DB 127 LSKIPDKNTLLKNIYKIEK---NTNINTADIEKTLITLGYEKLTVIIPGFTVKGEIID 183
QY 44 ----SESNEI-----SSKEIAKCSBPKPYLTKALQKLDKLLSKRSLODERTVIVYVT 95
DB 184 IYFPGEQNPRIALNFDKIEBKFNPLTQLKHDNEILEFQILPKKSIINDDKTINTLKT 243
QY 96 DTQKANIQLISELEE 110
DB 244 KIKSVEYKKILEELD 258

RESULT 15
PCT-US03-27401-259
; Sequence 259, Application PC/TUS0327401
; GENERAL INFORMATION:
; APPLICANT: TUFTS UNIVERSITY
; TITLE OF INVENTION: STREPTOCOCCUS PNEUMONIAE FOR DIAGNOSIS, TREATMENT AND
; TITLE OF INVENTION: PREVENTION OF ACTIVE INFECTION
; FILE REFERENCE: 700355-52941-PCT
; CURRENT APPLICATION NUMBER: PCT/US03/27401
; CURRENT FILING DATE: 2003-09-02
; PRIOR APPLICATION NUMBER: US 60/407,082
; PRIOR FILING DATE: 2002-08-30
; NUMBER OF SEQ ID NOS: 560
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 259
; LENGTH: 428

```

```

: TYPE: PRT
: ORGANISM: Streptococcus pneumoniae
PCT-US03-27401-259

Query Match      13.9%; Score 79.5; DB 1; Length 428;
Best Local Similarity 25.8%; Pred. No. 1.3e+02;
Matches 24; Conservative 20; Mismatches 36; Indels 13; Gaps 3;

QY 3 KINDINDLVNATFOVKKF-----FRDTKKKF---NLNYEEIYILNHLRSESNEISSK 52
Db 283 KADDWTDIVKTHAQSPDELVSYIKETLISFFGOYRMNENWVSVLEIGRDYOKELSLK 342

QY 53 ETAKCEPKPYLTALOKLD---LKLISKR 82
Db 343 DISKALFINPVYLGQLIKETDSTFABELNKOR 375

Search completed: November 17, 2003, 12:49:36
Job time : 68 secs
```